(9) EVIDENCE APPENDIX

Publications:

Lopez-Fuertes *et al.*, DNA vaccination with linear minimalistic (MIDGE) vectors confers protection against *Leishmania major* infection in mice, Vaccine 21 (2002) 247-257;

Kalderon *et al.*, A short amino acid sequence able to specify nuclear location, 1984, Cell 39, 499-509;

Kalderon *et al.*, Sequence requirements for nuclear location of simian virus 40 large-T antigen, 1984, Nature 377, 33-38;

Zanta *et al.*, Gene delivery: A single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus, 1999, Proc Natl Acad Sci USA, 96(1):91–6.

The listed references were referenced in a prior response submitted with the Declaration of Dr. Timon-Jiminez and with a response filed by applicant in a Request for Reconsideration filed May 28, 2009. They were cited for propositions that are discussed in this appeal.

ELSEVIER

Vaccine 21 (2002) 247–257

Vaccine

www.elsevier.com/locate/vaccine

DNA vaccination with linear minimalistic (MIDGE) vectors confers protection against *Leishmania major* infection in mice

L. López-Fuertes [a], E. Pérez-Jiménez [b], A.J. Vila-Coro [a], F. Sack [c], S. Moreno [a], S.A. Konig [a], C. Junghans [c], B. Wittig [c,d], M. Timón [a,*], M. Esteban [b]

[a] *Mologen Molecular Medicines S.L., Antonio de Cabezon 83, Piso 2, 28034 Fuencarral, Madrid, Spain*
[b] *Department of Molecular and Cellular Biology, Centro Nacional de Biotecnología, CSIC, UAM, E-28049 Madrid, Spain*
[c] *Mologen Holding AG, Berlin, Germany*
[d] *Department of Molecular Biology and Biochemistry, Fachbereich Humanmedizin, Freie Universitaet Berlin, Berlin, Germany*

Received 13 May 2002; received in revised form 14 August 2002; accepted 30 August 2002

Abstract

Immunization protocols based on priming with plasmid DNA and boosting with recombinants of vaccinia virus (rVV) encoding the same antigen offer great promise for the prevention and treatment of many parasitic and viral infections for which conventional vaccination has little or no effect. To overcome some of the potential problems associated to the use of plasmids, we have developed minimalistic, immunogenically defined, gene expression (MIDGE®) vectors. These linear vectors contain only the minimum sequence required for gene expression and can be chemically modified to increase the immune response. Here, we demonstrate that MIDGE vectors coding for the LACK antigen confer a highly effective protection against *Leishmania* infection in susceptible Balb/c mice. Protection is achieved at lower doses of vector compared to conventional plasmids. This efficacy could be greatly improved by the addition of a nuclear localization signal (NLS) peptide to the end of the MIDGE vector. In fact, immunization with two doses of NLS-modified MIDGE conferred similar or even better protection than that achieved by priming with plasmid DNA followed by boosting with rVV. These results demonstrate that MIDGE vectors are a good alternative to plasmid and rVV for immunization.

Keywords: MIDGE; DNA vaccine; *Leishmania*

1. Introduction

Leishmaniasis is a disease with different clinical manifestations produced by Trypanosomatidae of the genus *Leishmania*. Depending on the host immune response, the strain and the virulence of the parasite, the clinical manifestations extend from self-healing cutaneous lesions to the visceral form of the disease, the latter being fatal if untreated. Worldwide there are 2 million new *Leishmania* cases each year and the disease is endemic through parts of Africa, southern Europe and Central and South America [1]. With the advent of the HIV epidemic, leishmaniasis has surged as a reactivated infection in AIDS patients in many countries. Moreover, in Spain and the southwest of Europe leishmaniasis is zoonotic and dogs are the main reservoir host. Epidemiology of the disease revealed that between 10 and 37% of dogs in the Mediterranean area are infected and develop the visceral form of the disease [2].

Eradication of the disease has proven difficult. Chemotherapy has only a modest effect and there is no effective and safe vaccine against any form of clinical leishmaniasis. However, individuals who recovered naturally from infection develop strong immunity against re-infection suggesting that vaccination against leishmaniasis is feasible. Infection of inbred strains of mice with *Leishmania major* provides the best model for the immunoregulation that occurs during a cell mediated response to this intracellular pathogen [3]. Using the *Leishmania major* model, studies demonstrated that the generation of protective immunity against *L. major* is T cell and cytokine mediated [4]. Expansion of a Th1 subset of $CD4^+$ T lymphocytes secreting IFN-γ and IL-12 is associated with resistance to infection [4]. By contrast, susceptible mice expand $CD4^+$ T lymphocytes belonging to the Th2 subset, which secretes IL-4, IL-10 and IL-13 [5]. In human and dog, resistance to visceral leishmaniasis is also associated with the generation of a Th1 response

Abbreviations: MIDGE, minimalistic, linear expression vectors; NLS, nuclear localization signal

* Corresponding author. Tel.: +34-91-728-07-13; fax: +34-91-358-21-60.
E-mail address: timon@mologen.com (M. Timón).

PII: S0264-410X(02)00450-4

or a mixture of Th1 and Th2 cytokine patterns. Here, the absence of a Th1 response is associated with chronicity of the disease and high mortality in untreated cases [4]. Taken together these findings show that, as in the mouse model, the outcome of the disease appears to be determined and regulated by the balance between the two T cell populations [6,7].

Several antigens have been used in experimental vaccination trials in murine leishmaniasis, achieving various levels of protection. Among these are *L. major* gp63 [8], gp46 [9], p4, p8 [10] and LACK [11]. The LACK antigen is a 36 KD protein highly conserved among related *Leishmania* species and is expressed in both the promastigote and the amastigote forms of the parasite. LACK is a preferential target for the early anti-parasite immune response [12]. The infection induces a strong anti-LACK response and the early activated LACK reactive cells exhibit a Th2 phenotype [13]. Immune interventions aimed at modifying the T cell repertoire may be used to alter the course of an infection. The option to modulate this rapid anti-LACK immune response by exogenous IL-12 or IFN-γ suggests that these cells are not yet fully mature Th2 cells and so can be redirected [13]. Furthermore, vaccination trials with soluble leishmania antigens or a single LACK protein in the presence of recombinant IL-12 have been shown to induce a protective Th1 response [13,14]. Thus, it seems that deviation to Th1 of the strong and specific Th2 immune response against LACK in Balb/c mice induces resistance to infection [15].

During the last decade, DNA immunization has been promoted as a new alternative for achieving specific immune responses. Protective responses with DNA vaccines against several pathogens have been demonstrated [16,17]. The ability of plasmid DNA encoding specific antigens to induce both $CD4^+$ and $CD8^+$ T cell responses could be of particular use for protection against diseases that require cell mediated immunity, including leishmaniasis. The low production costs and the high structural stability of DNA make DNA vaccination a very attractive tool for immunization. On the other hand, recombinants of vaccinia virus (rVV) expressing different foreign antigens have been successfully used to elicit protective immunity to a variety of pathogens [18–21]. For instance we have shown that immunization with rVV expressing the *L. amazonensis* gp46 antigen elicits significant protection and long-term immunological memory in BALB/c mice [22]. Experiments in mice and other models have shown that a combination of these two approaches (priming with DNA and boosting with recombinant vaccinia expressing the same protein) is associated with the highest immunogenicity and protective efficacy against several infectious agents. Using this approach we have demonstrated a good protection against *L. major* in mice with the LACK antigen [23].

Expression plasmids typically used in DNA-based vaccination usually contain a transcription unit and bacterial sequences necessary for plasmid amplification and selection. In order to avoid some of the potential and principal problems associated with the use of these plasmids, we have designed minimalistic, immunogenically defined gene expression (MIDGE) vectors that contain only the eukaryotic gene expression cassette [24]. Thus, MIDGE constructs carry no sequence elements other than those needed for gene transfer. Their ends provide unique and selective target sites for the coupling of modifying molecules. To assess their utility for vaccination against leishmaniasis, MIDGE vectors expressing the LACK antigen from *L. infantum* were constructed. A nuclear localization signal (NLS) peptide derived from the T antigen of the SV40 virus was covalently linked to modification target sites at the ends of the LACK expressing MIDGE vectors. NLS peptides conjugated to plasmids have been shown to increase transfection efficiency which could result in enhanced immunogenicity [25–27]. NLS-modified and non-modified MIDGE vectors were investigated in their ability to confer protection against challenge infections, and compare in this aspect to various schemes of plasmid-based and rVV vaccinations.

Here, we show that immunization of mice with MIDGE vectors coding for p36/LACK, induced the same level of protection against challenge with *L. major* than equimolar amounts of plasmid coding for the same antigen. Most importantly, two doses of NLS-modified MIDGE conferred full protection against the same antigen, resembling the results obtained by priming with plasmid DNA and boosting with rVV, both expressing LACK.

2. Material and methods

2.1. Construction of plasmids, MIDGEs and modified MIDGEs

The cDNA encoding the LACK protein from *L. infantum* was obtained as described previously [11] and inserted downstream of the CMV promoter in the *Eco*RV site of the pcDNA3.1 expression vector (Invitrogen, San Diego, CA). The coding sequence of LACK/p36 was subcloned into the *Sst*I and *Kpn*I sites of the pMOK plasmid (Mologen, Berlin, Germany) generating the plasmid pMOK-p36. The MIDGE-p36 construct was derived from the pMOK-p36 plasmid after complete digestion with *Eco31I* (MBI Fermentas, Vilnius, Lithuania). The ends were used to ligate hairpin oligodeoxyribonucleotides (ODNs) with T4 DNA ligase. The mixture was concentrated and treated with *Eco31I* and T7 DNA polymerase in the absence of deoxyribonucleotides. The DNA was purified by anionic exchange column chromatography (Merck EMD-DMAE, sodium phosphate pH 7.0–0.1 M NaCl).

The NLS peptide (PKKKRKVEDPYC, a generous gift from Dr. P. Henklein, Charité, Berlin, Germany) was coupled to one of the hairpin ODN in two steps as previously described [24]. The resulting NLS-coupled ODN was purified. The plasmid pMOK encoding the surface antigen of the hepatitis B virus (HBsAg) was used as a control.

2.2. Construction of recombinant vaccinia viruses (rVV)

The cDNA encoding p36 from *L. infantum* was cloned into the pSC11 VV insertion plasmid under control of the p7.5 early/late viral promoter in the thymidine kinase (TK) locus. This plasmid contains the *E.coli* β-galactosidase gene under the control of the p11 viral late promoter (pSCp36). Vaccinia virus (VV) recombinants were derived from the wild type Western Reserve (WR) strain, and rVVp36 was prepared by transfecting with the insertion plasmid WR-infected BSC-40 cells [20]. The recombinant viruses were harvested 48–72 h post-infection and selected after plaque assay by addition of X-Gal to the agar. β-Galactosidase-producing plaques picked and re-plated three times and amplified following standard procedures [23,28].

2.3. Cells

African green monkey cells (BSC-40) and HeLa cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% newborn calf serum (NCS, Gibco BRL, Paisley, UK). Viruses were grown in HeLa and titrated in BSC-40 cells. COS-7 cells were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum (FCS, Gibco BRL, Paisley, UK).

2.4. Mice

Female 6–8-week-old BALB/c mice, maintained under pathogen-free conditions, were obtained from the facilities at the Centro Nacional de Biotecnología (Madrid).

2.5. Parasites and reagents

L. major (WHOM/IR/-173) was a kind gift from Dr. N. Glaichenhaus (CNRS, Valbonne, France). Promastigotes were cultured at 27 °C in Schneiders medium (Gibco BRL, UK) supplemented with 20% fetal calf serum and antibiotics. Parasites were expanded in BALB/c mice and *L. major* amastigotes were obtained from a hind footpad lesion. After transformation from amastigotes to promastigotes, a synchronized culture was established for 9 days until parasites were in a late stationary phase.

2.6. LACK/ p36 expression in transiently transfected cells

3×10^5 COS-7 cells were transiently transfected with equimolar amounts of the different DNA vectors (pMOK-p36: 4 μg; MIDGE-p36 and MIDGE-NLS-p36: 2.2 μg) using Lipofectamine transfection reagent (Gibco BRL, UK). One day later expression of LACK/p36 was detected by Western blot analysis of whole cell lysates. The Western blot was reacted with an anti-p36 rabbit polyclonal antibody as previously described (23). For band quantification, the Western blot was digitized and analyzed with the NIHimage program (National Institutes of Health, Bethesda Maryland). After equalization of the image, the mean density was measured for each band using a selection of fixed area and shape.

2.7. Immunization and infectious challenge

Groups of BALB/c mice (n = 10–11) were primed intradermally (i.d.) in the back with 100 μg per mouse of pMOK-p36 or pMOK-HBsAg in 100 μl volume. The groups primed with MIDGE-p36 and MIDGE-p36-NLS received 54.8 μg per mouse which is an equimolar concentration of the plasmids (Table 1). Two weeks later (14 d.p.i.) mice were boosted either i.d. in the back with the same amount of DNA or intraperitoneally (i.p.) with rVVp36 (5×10^7 pfu per mouse). The non-immunized group received the same volume of buffer (150 mM sodium phosphate). Three weeks after boosting (34 d.p.i.), three mice per group were sacrificed and sera were obtained. The following day (35 d.p.i.), mice (n = 7–8) were challenged subcutaneously in the right hind footpad with 5×10^4 live stationary phase *L. major* promastigotes. Lesion development at the inoculation site was measured weekly with a digital caliper (Mauser Digital, Switzerland) and expressed as the increase in thickness of infected versus uninfected hind foot. Mice were sacrificed at week 8 post-challenge and serum, lymph nodes and spleen collected.

2.8. Evaluation of cytokine production

Cytokine levels in cell culture supernatants were determined by ELISA. Single cell preparations from spleens were plated in triplicate at 4×10^6 cells/ml in 24 well plates (Nunc, Denmark). Soluble antigenic peptide (2 μg/ml) (a kind gift from Dr. N. Glaichenhaus [29]), p36 protein (2 μg/ml) prepared as previously described [11], soluble leishmanial antigen (4 μg/ml) and ConA (2 μg/ml) were added in a final volume of 2 ml/well. Supernatants were harvested after 48 and 72 h and stored at −80 °C until used. IFN-γ and IL-4 levels were assessed by specific ELISA using capture and

Table 1
Immunization regimes in mice

Group	Prime[a]	Boost[a]
1	pMOK-p36 (100 μg)	pMOK-p36 (100 μg)
2	MIDGE-p36 (54.8 μg)	MIDGE-p36 (54.8 μg)
3	MIDGE-p36-NLS (54.8 μg)	MIDGE-p36-NLS (54.8 μg)
4	pMOK-HBsAg (100 μg)	pMOK-HBsAg (100 μg)
5	pMOK-p36 (100 μg)	rVVp36 (5×10^7 pfu)
6	MIDGE-p36 (54.8 μg)	rVVp36 (5×10^7 pfu)
7	MIDGE-p36-NLS (54.8 μg)	rVVp36 (5×10^7 pfu)
8	Buffer[b]	Buffer[b]

[a] Injections were intradermic (i.d.) for DNA or intraperitoneal (i.p.) for vaccinia virus. Booster injections were given 2 weeks after priming.
[b] Mice from this group received 2 i.d. doses of 150 mM sodium phosphate buffer.

secondary antibodies from Pharmingen (Becton Dickinson, USA) following the manufacturer's instructions.

2.9. Detection of total IgG antibodies against p36

One day before challenge and 7 weeks post-challenge, serum was collected from each group of animals and specific anti-p36 IgG antibodies and their isotypes (IgG1 and IgG2a) were analyzed by ELISA. In brief, 96-well Maxisorp plates (Nunc, Denmark) were coated overnight at 4 °C with recombinant p36 (5 μg/ml) prepared as previously described [11]. Serum samples were diluted 5- and 10-fold in blocking buffer (1% BSA in PBS-T), added in 50 μl/well and incubated 1 h at 37 °C. Peroxidase-conjugated goat anti-mouse total IgG, IgG1 or IgG2a (Southern Biotechnology Associated, Birmingham, AL) was added and incubated for 1 h at 37 °C. Plates were then reacted with peroxidase substrate OPD (Sigma, St Louis, MO) and absorbance was read at 492 nm on a Labsystem Multiskan Plus plate reader (Tecan Maguellan, Sunrises).

3. Results

3.1. LACK/p36 encoding MIDGE vectors

The p36 encoding sequence was inserted into pMOK generating the plasmid pMOK-p36 (Fig. 1A). In this vector the LACK/p36 protein is expressed under the control of promoter and enhancer sequences from the immediate early region of HCMV. LACK/p36 expression was improved in the pMOK vector by introducing an intron sequence between the CMV promoter and the LACK/p36-encoding sequence. pMOK-p36 contains bacterial sequences required for plasmid amplification and selection (Kan^R).

The MIDGE-p36 construct was generated from the plasmid pMOK-p36 as described in Section 2 (Fig. 1A).

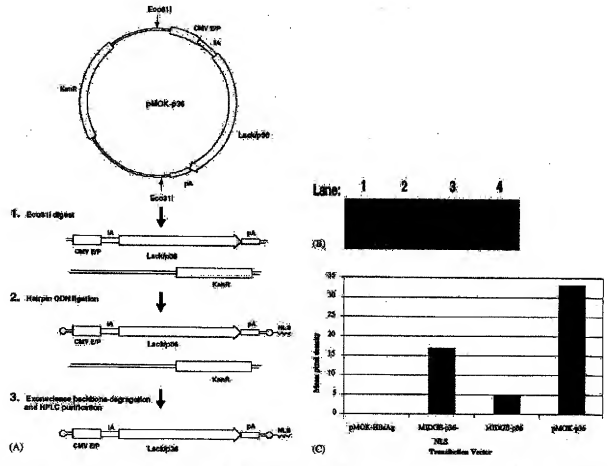

Fig. 1. Construction of MIDGE vectors and expression of LACK/p36 from different constructs. (A) The LACK-encoding pMOK plasmid was used to generate the LACK-encoding MIDGE constructs. By using NLS-modified hairpin ODN we generated the MIDGE-p36-NLS construct; (B) Expression of LACK/p36 protein in transiently transfected COS-7 cells. Cells were transiently transfected with equimolar amounts of DNA (pMOK-p36 and pMOK-HbsAg: 4 μg; MIDGE-p36 and MIDGE-p36-NLS: 2.2 μg) using lipofectamine transfection reagent. The expression of the protein was detected by Western blot analysis of lysate cells. One representative experiment is shown. Lane 1: pMOK-HBsAg; lane 2: MIDGE-p36-NLS; lane 3: MIDGE-p36; lane 4: pMOK-p36; (C) The figure shows the result of band densitometry expressed in arbitrary units (see Section 2).

MIDGE-p36 is a linear construct with covalently-closed ends. This expression unit contains only the promoter, intron, LACK/p36 encoding, and polyadenylation sequences (Fig. 1A). The linearity of the construct was confirmed by digestion with an appropriate restriction enzyme that has a single recognition site in the construct and generated two DNA fragments of the expected size (data not shown). In order to increase transfection efficiency and the expression of the antigen, we linked to the LACK expressing MIDGE vector a nuclear localization sequence signal (NLS) derived from the T antigen of the SV40 virus. The NLS peptide was coupled to one of the hairpin ODN in two steps as previously described [24].

3.2. Efficient LACK/p36 expression in cells transiently transfected with LACK/p36-encoding plasmid, MIDGE and MIDGE–NLS

Transient expression of p36 from plasmid pMOK-p36, the corresponding MIDGE construct and MIDGE-modified with the NLS sequence was tested in vitro in COS-7 cells transfected with DNA. The transfection was done using equimolar concentrations of each construct and the level of LACK/p36 expression was measured in lysed cells (Fig. 1B). The MIDGE construct expressed about six-fold less LACK/p36 protein than the plasmid (Fig. 1B and C). By contrast, NLS-modified MIDGE expressed a higher level of protein than unmodified MIDGE, but two-fold less than plasmid (Fig. 1C).

3.3. Efficacy of MIDGE vectors to confer protection against leishmaniasis

The efficacy of MIDGE vectors to trigger a protective immune response was tested in a murine model of *Leishmania* infection. For these studies, we compared protocols based on the use of a conventional plasmid and simple or NLS-modified MIDGE vectors (Table 1). Mice were primed with different constructs encoding the p36 antigen from *L. infantum*. Fourteen days later, half of the groups were boosted with rVVp36 and the others with the three different DNA vectors (Table 1). Three weeks after boosting, all animals were challenged with 5×10^4 late stationary phase *L. major* promastigotes in the right hind footpad (see Section 2). Progression and size of the lesion were measured weekly. Mice were sacrificed 8 weeks after challenge because of the large size of the lesion in control groups.

3.3.1. Priming/boosting with MIDGE vectors confers similar protection against L. major challenge than priming/boosting with a classical plasmid

All immunized groups showed several degrees of protection along the 8 weeks following the challenge (Fig. 2A). The protocol based on priming and boosting with MIDGE vector induced the same level of protection than that observed with the protocol based on priming and boosting with plasmid (Fig. 2A). Thus, both groups presented a reduction in lesion size of approximately 45% 8 weeks after challenge (Table 2).

3.3.2. Priming/boosting with MIDGE-p36-NLS confers high protection against L. major infection

When we compared the above protocols with priming/boosting with MIDGE–NLS, we observed a higher reduction in lesion size with the latter protocol (Fig. 2A). The overall protection was almost 80% compared to the control group immunized with plasmids coding for an irrelevant antigen (Mann–Whitney test, $P = 0.0056$) (Table 2). By following the progression of the disease weekly, both MIDGE and MIDGE–NLS showed a good level of

Table 2
Specific IgG2a/IgG1 ratios pre- and post-challenge, lesion size and lymph node weight at week 8 after challenge

Groups	Pre-challenge	Post-challenge		
	IgG2a/IgG1[a]	Lesion size[b] (%)	Lymph node weight[c] (mg)	IgG2a/IgG[d]
pMOK-p36/pMOK-p36	–	55	70 ± 17	0.9
MIDGE-p36/MIDGE-p36	–	56	75 ± 07	1.0
MIDGE-p36-NLS/MIDGE-p36-NLS	–	21*	54 ± 32	1.3[d]
pMOK-HBsAg/pMOK-HBsAg	–	100	74 ± 14	1.0
pMOK-p36/rVVp36	4.25	31**	45 ± 19***	1.66[d]
MIDGE-p36/rVVp36	5.76	62	93 ± 29	1.0
MIDGE-p36-NLS/rVVp36	3.98	47	71 ± 20	1.3[d]
Buffer/buffer[e]	–	85	73 ± 23	0.96
Positive control[f]	0.5			

[a] IgG2a/IgG1 represents the ratio of the mean absorbance values at 492 nm of anti-LACK specific antibodies from each immunization group determined by an indirect ELISA.

[b] Percentage of lesion size development relative to the lesion size in the group inoculated with two doses of pMOK-HBsAg at 8 weeks after challenge.

[c] Eight weeks after challenge 7–8 mice were sacrificed and each draining lymph node was extracted and weighted. Means ± S.D. are shown.

[d] Groups of mice with an IgG2a/IgG1 ratio > 1 at week 8 post-challenge are underlined.

[e] These mice received two doses of 150 mM sodium phosphate.

[f] This represents the IgG2a/IgG1 ratio in a non-immunized mouse infected with 5×10^4 live stationary phase *L. major* promastigotes.

* $P = 0.0056$.

** $P = 0.0042$.

*** $P = 0.0023$.

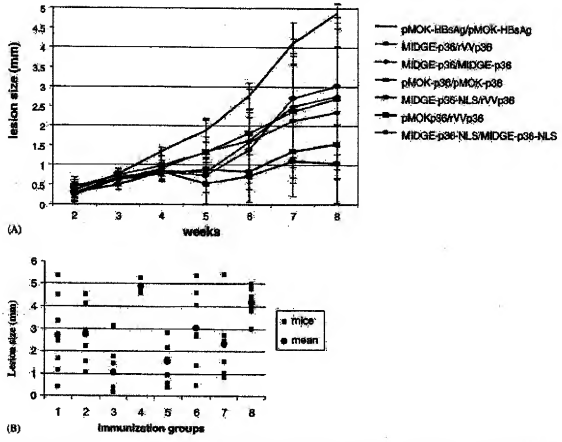

Fig. 2. Course of *L. major* infection in mice. (A) To examine the protective efficacy of the vectors, 3 weeks after boosting, groups of 7–8 mice were challenged s.c. in the right hind footpad with 5×10^4 live stationary phase *L. major* promastigotes. The state of the infection was followed weekly. Lesion size was determined by measuring the thickness of the right footpad with a digital caliper, and expressed as the difference of thickness between the infected and the uninfected collateral footpads. The figure shows the mean values ± S.D. of lesion size in each immunization group measured at weekly intervals; (B) Lesion development at 8 weeks post-challenge. Black squares represent the lesion score for each individual mouse and grey dots the mean value for each immunized group.

protection up to week 5 post-challenge (Fig. 2A). However, after this time point, the lesion developed in the group primed/boosted with MIDGE-p36 while it did not progress in the group primed/boosted with MIDGE-p36-NLS. This was manifested after examination of lesion size in individual animals (Fig. 2B). In fact, 57% (four out of seven) of the animals from this group had virtually no lesion (<0.5 mm) 8 weeks after challenge.

3.3.3. The protocol based on priming/boosting with MIDGE-p36-NLS induces at least as good protection against L. major than priming with plasmid DNA vector and boosting with rVVp36

Although vaccination with DNA alone has proven effective in animal models, it has been recently shown that a protocol based on priming with DNA vector and boosting with vaccinia virus recombinants expressing different genes induces high protection upon challenge with several pathogens including *Leishmania* [20,23,30–36]. We, therefore, compared the efficacy of priming/boosting between classical plasmid and MIDGE vectors with rVV expressing p36. As shown in Fig. 2A, the extend of protection triggered by the protocol based on priming/boosting with MIDGE-p36-NLS was similar to that induced by priming/boosting with pMOK-p36 and rVVp36, showing not statistically significant differences between both groups. At the end of the experiment, the average reduction in lesion size was almost 80% in the group immunized with two doses of MIDGE-p36-NLS (Table 2), whereas the group immunized with pMOK-p36/rVVp36 presented an average reduction of 64% compared to the group immunized with plasmids coding for an irrelevant antigen (Mann–Whitney test, $P = 0.0056$ and $P = 0.0042$, respectively) (Table 2). The difference in lesion size between MIDGE-p36-NLS/MIDGE-p36-NLS and pMOK-p36/rVVp36 immunized animals was not significant (Mann–Whitney test, $P = 0.22$).

When used only for priming, both MIDGE and MIDGE–NLS showed a good level of protection up to week 5 post-challenge (Fig. 2A). However, after this time point,

the lesion developed very rapidly in the group primed with MIDGE-p36 and more progressively in the one primed with MIDGE-p36-NLS (Fig. 2A), showing a reduction in lesion size of 38 and 53%, respectively, 8 weeks after challenge as compared to the group immunized with two doses of pMOK-HBsAg (Mann–Whitney test, $P = 0.52$) (Table 2).

3.4. Reduction in lesion size correlates with small lymph nodes

Reduction in lesion size indicates that parasite replication is diminished in the local lymph node. To test that, each popliteal lymph node from the infected leg was extracted and weighted. Average lymph node weight was significantly lower in the two most protected groups (Table 2). In the group immunized twice with MIDGE-p36-NLS the average lymph node weight was almost 30% lower than that in controls and 40% lower in the group immunized with pMOK-p36 and rVVp36 (Mann–Whitney test, $P = 0.25$ and 0.0023, respectively).

Interestingly, the group immunized with MIDGE-p36/rVVp36 had the largest lymph node of all, even more than controls. This may reflect the rapid growth of the lesion in this group over the last 3 weeks of the analysis (Fig. 2A). Alternatively, a large lymph node could reflect an active lymphocyte proliferation. This possibility could not be ruled out directly due to low number of cells recovered from the local lymph node. However, phenotypic analyses of splenocytes from all groups showed no differences in the percentage of any cell subset or activation market (data not shown). These suggested the lack of a general immune activation.

3.5. Humoral immune responses before L. major challenge

Since the outcome of the disease may be determined by the extent and the type of the immune response, we decided to check the antibody response before the challenge. Three weeks after boosting, serum was collected from each group of animals and the specific anti-p36 immunoglobulin G levels and their isotypes (IgG1 and IgG2a) were evaluated. As positive control, two sera from animals infected with *L. major* were used. Specific anti-p36 total IgG was only detected in sera from animals immunized with protocols based on priming with DNAp36 (pMOK, MIDGE or MIDGE–NLS) and boosting with rVVp36 (Fig. 3). The main isotype produced was IgG2a indicating a prevalent Th1 immune response (Table 2). By contrast, control animals developed a characteristic Th2 immune response (Table 2).

3.6. Humoral and cellular immune responses after L. major challenge

Since activation of a Th1 type of immune response may be required to induce protection against *Leishmania* infection, we were interested in characterizing the changes in immunoglobulin isotypes and cytokine production following challenge in animals. Spleens and sera were taken from sacrificed mice and specific IgG antibodies, and IFN-γ and IL-4 production were evaluated. Analysis of specific IgG levels and their isotypes revealed that all groups had developed anti-p36 total IgG as expected because all of them had been challenged with the parasite. The levels were slightly higher in the groups primed and boosted with MIDGE-p36 and MIDGE-p36-NLS (Fig. 3). Most importantly, the highest overall IgG2a/IgG1 ratios were present in the groups that showed the highest levels of protection, indicating a prevalence of a Th1 immune response (Table 2).

In order to further correlate the generation of a Th1 response with protection, pooled splenocytes from each group were stimulated with either recombinant p36 protein, an immunodominant class II-specific peptide from this protein [29], a lysate of *Leishmania* parasite or Concavalin A as positive control. IFN-γ (a Th1 cytokine) and IL-4 (Th2) production was determined in the supernatants 72 h after stimulation by ELISA. No IL-4 could be detected in any of the

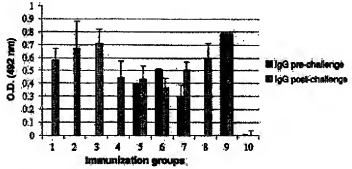

Fig. 3. Specific IgG antibodies before and after challenge with *L. major*. Sera were collected 1 day before challenge with *L. major* parasite and 8 weeks post-challenge. Specific anti-p36 IgG antibodies were analysed by ELISA. Black bars represent the mean ± S.D. of three mice per group (pre-challenge) and grey bars represent the mean ± S.D. of 7–8 mice per group (post-challenge). Groups 1–8 are defined in Table 1. Group 9 represents serum from a non-immunized animal infected with 5×10^4 live stationary phase *L. major* promastigotes (the mean ± S.D. of triplicate wells is presented). Group 10 represents serum from a naive mouse (the mean ± S.D. of triplicate wells is presented). Serum from group 9 was used to normalize both ELISAs.

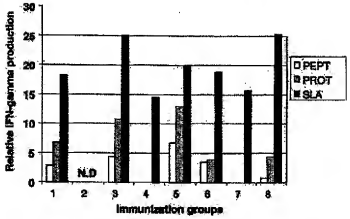

Fig. 4. IFN-γ production by in vitro stimulated splenocytes. Eight weeks after challenge, spleens were removed and pooled in each group. Cells were stimulated in vitro with recombinant LACK protein, the main immunogenic peptide from LACK [14], soluble leishmanian antigen (SLA) or ConA. Seventy-two hours post-stimulation supernatants were harvested and IFN-γ production was assessed by specific ELISA. The concentration of IFN-γ was calculated by extrapolating the absorbance value to a standard curve made with known concentrations. Values between groups were standardized by using the IFN-γ production after stimulation with ConA as the maximum value. Bars represent the percentage of IFN-γ production relative to the production after ConA stimulation. Groups 1–8 are those defined in Table 1. ND: not determined.

groups presumably due to the low sensitivity of the assay. By contrast, relatively high levels of IFN-γ were observed in all groups after stimulation with LSA. Stimulation with the p36 protein or the peptide induced a higher IFN-γ production in the groups showing the best protection (Fig. 4).

4. Discussion

DNA immunization represents a novel and interesting approach to vaccine development and immunotherapy, particularly for applications where cell-mediated immune responses are required. Several obstacles are hampering the translation of this promising technology to the clinic, including safety concerns associated to the use of conventional plasmids and the high doses required to obtain protective immunization. We and others have demonstrated that the use of plasmid DNA as priming agent followed by boosting with recombinants of vaccinia virus coding for the same antigen represents a very effective protocol for immunization and protection [20,23,30–36]. Although this strategy allows a considerable reduction in the amount of DNA required, the use of recombinant viruses abrogates some of the advantages of DNA vaccination. In the present work we have studied the ability of minimalistic, immunogenically-defined, gene expression vectors to induce protection in a mouse model of *Leishmania* infection. MIDGE vectors offer several significant advantages over conventional plasmids, such as small size, absence of antibiotic resistance genes and selective target sites for chemical linkage of peptides, proteins, sugars, etc. [24]. The efficacy of MIDGE vectors was compared to that of plasmids in several immunization protocols including boosting or not with rVV (Table 1).

The very conserved LACK antigen from *L. infantum* was cloned into both a plain MIDGE vector and a MIDGE-modified by covalently-linked nuclear localization signals (NLS) (Fig. 1A). NLS has been shown to increase the translocation of both proteins [37] and DNA [25–27] from the cytoplasm to the nucleus thus bypassing one of the main barriers for the expression of foreign DNA. To test whether this increase in nuclear transport correlated with a higher antigen expression, COS-7 cells were transfected in vitro and LACK expression was detected by Western blotting. As shown in Fig. 1B, the addition of one copy of NLS resulted in a three-fold increase in antigen expression compared to plain MIDGE. However, this expression was still lower than that achieved by transfection with an equimolar amount of plasmid coding for the same protein, indicating that increased transport to the nucleus is not sufficient to achieve high antigen expression.

MIDGE vectors expressing the LACK protein were then used for in vivo immunization of Balb/c mice and their efficacy in inducing protection was determined by challenging the mice with promastigotes from *L. major*. This is a well established model that has been extensively used to study the immunological parameters involved in *Leishmania* infection and protection as well as for the validation of vaccine candidates against this parasite [3,6]. As shown in Fig. 2, two doses of plain MIDGE-p36 confer partial but sustained protection against the challenge. This protection is identical to that induced by two doses of plasmid, but it is achieved with only half the amount of DNA (Table 1). The development of the lesion also follows the same kinetic in both groups of immunized animals (Fig. 2A) suggesting that these vectors induce both quantitatively and qualitatively similar immune responses in spite of the lower antigen expression observed in MIDGE-transfected cells in vitro

(Fig. 1B). Most importantly, the best protection against challenge was observed in the group immunized with two doses of NLS-modified MIDGE vector (Fig. 2A). In this group, four out of seven mice were completely protected at the termination of the study and only one presented a lesion of considerable size (Fig. 2B). These results were even better than those obtained by priming with plasmid and boosting with rVV, which is considered to be one of the best immunization protocols available [36,38]. In addition, both MIDGE-p36 and MIDGE-p36-NLS conferred a good but short protection against challenge with *L. major* when used as priming vectors followed by boosting with rVVp36 (Fig. 2). Mice from these groups were completely protected up to 5 weeks post-challenge but from this time the lesion developed very rapidly, particularly in the group primed with MIDGE-p36 (Fig. 2A).

Trying to understand the immunological mechanisms underlying these results, we analyzed serum from three mice in each group for specific anti-LACK IgG antibodies and their isotypes at the time of challenge. It has been previously shown by many authors that in this model of *Leishmania* infection, there is a clear correlation between resistance to infection and the development of a Th1 type of response, whereas susceptibility correlates with the development of a Th2 response [3,5]. Although protection against *Leishmania* is not mediated by antibodies [7], IgG isotypes are a valid downstream indicator of the type of T-helper immune response generated. Specific IgG antibodies were only detected in those groups receiving rVVp36 at the time of boosting and not in those receiving two doses of DNA (Fig. 3). The explanation for these results could be that LACK is a cytoplasmic protein that cannot be detected by B cells when expressed from a DNA vector. By contrast, Vaccinia virus induces lysis of infected cells [21] and so the whole protein becomes exposed to the immune system. The IgG antibodies produced by mice boosted with rVVp36 were predominantly of the IgG2a isotype (Table 2), indicating the induction of a Th1 response, as is normally the case when using rVV [21].

In contrast to these results, all mice had specific anti-LACK antibodies 8 weeks after challenge (Fig. 3), implying that replication of the parasite and cell lysis had occurred in all of them. This was also suggested by the increase in size of the local lymph node at the termination of the experiment, which was particularly important in those groups with the lowest protection (Table 2) and which could not be attributed to a higher immune response (data not shown). Interestingly, the groups immunized with two doses of any DNA vector had the highest antibody levels. This suggests that, although undetectable at the time of the analysis, DNA immunization had probably induced a low and/or transient production of antibodies that was boosted by the infection. However, as mentioned above, antibodies do not seem to be involved in protection because their level before or after challenge did not correlate with lesion development. Specific IgG isotypes were also determined at the time of sacrifice. As shown in Table 2, those groups with a higher IgG2a/IgG1 ratio (indicating a bias towards a Th1 response) were the best protected at week 8 post-challenge. To further correlate protection with the generation of a Th1 response, pooled splenocytes from every group of mice were re-stimulated in vitro with different stimuli and the IFN-γ production (a Th1 cytokine) was measured in the supernatants. Results were not very conclusive, probably because by week 8 after challenge the immune system had time to generate a response against many antigens from the parasite. As a consequence, the specific anti-LACK cellular response, important in the early stages of the infection, may be difficult to detect. This is demonstrated by the fact that splenocytes from most groups were able to produce INF-γ in response to a whole *Leishmania* lysate (Fig. 4). Nonetheless, it seems that only splenocytes from groups showing the best protection (group 3: MIDGE-p36-NLS/MIDGE-p36-NLS and group 5: pMOK-p36/rVVp36) produced a good amount of IFN-γ in response to LACK (Fig. 4). More experiments have to be done to confirm this correlation, preferably soon after the challenge because the anti-LACK response normally occurs very early after infection and so it may be at this point when generation of the right (Th1) response is critical for controlling replication of the parasite.

Altogether, we have shown that MIDGE vectors are very efficient at inducing protection against challenge in a well-studied murine model of *Leishmania* infection. Although in vitro, the plain MIDGE vector induced lower protein expression than plasmid, in vivo they both conferred similar protection. This suggests that antigen load may not be the only limiting factor when vaccinating with DNA and that MIDGE vectors have an unknown feature that makes them immunologically more efficient. Alternatively, it could be that MIDGE vectors induce higher protein expression in vivo or that the amount of antigen expressed after two doses is above the threshold required for the induction of an immune response.

When given only for priming, half the amount of both plain MIDGE and NLS-modified MIDGE vectors in combination with rVV conferred almost full protection during the first 5 weeks post-infection. After this time, however, the parasite took over and the lesion developed very quickly in these two groups. By contrast, when plasmid was used for priming, the infection was controlled over a long period of time. It is not clear why priming with plasmid DNA followed by boosting with recombinant vaccinia virus is such an effective protocol for immunization whereas both used in the opposite order are not effective. One of the most plausible explanations is that rVV given first strongly primes many anti-viral T and/or B cells and so dilute the antigen-specific ones, which are then very difficult to re-amplify by the DNA booster [21–40]. When DNA is given first, the response maybe low but specific against the cloned antigen and so can be easily amplified by the very immunogenic rVV [40]. Our results indicate that to achieve a strong and long-lasting immune response with this protocol it may be necessary to generate enough antigen during the priming step. This would

explain why plasmid was better than MIDGE vectors as priming agent and why the group primed with NLS-modified MIDGE contained the infection slightly better than the one primed with plain MIDGE. Experiments are underway to test this possibility by using larger amounts of MIDGE vectors to achieve similar antigen load than when priming with plasmid.

The most important outcome from this work was the strong protection obtained with only two doses of NLS-modified MIDGE coding for p36. To our knowledge, this is the first time that an immunization protocol based only on DNA gives comparable or even better protection than one based on plasmid DNA + rVV. This result can not be attributed to a higher antigen load induced by the MIDGE vector system because, as already emphasized, plasmid was better for protein expression at least in vitro (see above). Most likely, the presence of the NLS peptide covalently linked to the DNA is making the vector qualitatively different. In fact, when working on a different model, we have observed that immunization with NLS-modified MIDGE vectors predominantly induce a Th1 type of response in mice, even in conditions where other vectors induce a Th2 response (Moreno et al., manuscript in preparation). The basic mechanism underlying this bias is still unclear, but it could explain the good protection conferred against *Leishmania* infection, reported to be highly dependent on the generation of a Th1 response [3]. Although the design of the present experiment, more focused on protection than on basic immunological studies, does not allow for a more precise determination of the type of immune response generated, the data available suggest that the most protected groups had more Th1 response. On the basis of all the data, we can conclude that a long-lasting protection against *Leishmania* infection from immunization with the LACK antigen first requires a certain amount of antigen to trigger the specific immune response. Once generated, only if the response is of the Th1 type, the infection will be controlled.

In summary, the evidence presented here shows that MIDGE vectors are a true alternative to plasmids and recombinant viruses for protective immunization. MIDGE vectors offer great hope for the development of effective and safe vaccines against *Leishmania* and other devastating infections for which conventional vaccines have no effect.

Acknowledgements

We thank Dr. V. Larraga for the generous gift of the plasmid containing the *p36* gene from *L. infantum*, Dr. P. Henklein for the NLS peptide and Dr. Nicholas Glaichenhaus for providing *L. major* strain. We thank Dr. Matthias Schroff and Detlef Oswald for the cloning, and Jens Alfken and Gaby Glowacz for the production of the plasmids and MIDGE vectors. We also thank Dr. Colin Smith for the critical reading of the manuscript. This investigation was supported, in part, by a grant from Comunidad Autónoma de Madrid (0.8.2/0057/2001) to M.E.

References

[1] World Health Organization, WHO. Technical Report Series, No. 793, 1990.
[2] Bettini S, Gradoni L. Canine leishmaniasis in the Mediterranean area and its implications for human leishmaniasis. Insect Sci Appl 1986;7:241–5.
[3] Reiner SL, Locksley RM. The regulation of immunity to *Leishmania major*. Ann Rev Immunol 1995;13:151.
[4] Mattner F, Magram J, Ferrante J, et al. Genetically resistant mice lacking interleukin-12 are susceptible to infection with *Leishmania major* and mount a polarized Th2 cell response. Eur J Immunol 1996;26:1553–9.
[5] Heinzel FP, Sadick MD, Mutha SS, Locksley RM. Production of interferon gamma, interleukin 2, interleukin 4, and interleukin 10 by CD4+ lymphocytes in vivo during healing and progressive murine leishmaniasis. Proc Natl Acad Sci USA 1991;88(16):7011–5.
[6] Scott P, Natovitz P, Coffman RL, Pearce E, Sher A. Immunoregulation of cutaneous leishmaniasis T cell lines that transfer protective immunity and exacerbation belongs to different T helper subsets and response to distinct parasite antigens. J Exp Med 1988;168:1675–84.
[7] Sher A, Coffman RL. Regulation of immunity to parasites by T cells and T cell-derived cytokines. Ann Rev Immunol 1992;10:385–409.
[8] Xu D, Liew FY. Genetic vaccination against leishmaniasis. Vaccine 1994;12:1534–6.
[9] Champsi J, McMahon-Pratt D. Membrane glycoprotein M-2 protects against *L. amazonensis* infection. Infect Immunol 1988;56:3272–9.
[10] Soong L, Duboise SM, Kima P, McMahon-Pratt D. *Leishmania pifanoi* amastigote antigens protects mice against cutaneous leishmaniasis. Infect Immunol 1995;63:3556–9.
[11] González-Aseguinolaza G, Taladriz S, Marquet A, Larraga V. Molecular cloning. Eur J Biochem 1999;259:909–16.
[12] Maillard I, Launois P, Himmelrich H, et al. Functional plasticity of the LACK-reactive Vbeta4-Valpha8 CD4(+) T cells normally producing the early IL-4 instructing Th2 cell development and susceptibility to *Leishmania major* in BALB/c mice. Eur J Immunol 2001;31(4):1288–96.
[13] Afonso LC, Scharton TM, Vieira LQ, Wysocka M, Trinchieri G, Scott P. The adjuvant effect of interleukin-12 in a vaccine against *Leishmania major*. Science 1994;263(5144):235–7.
[14] Mougneau E, Altare F, Wakil AE, et al. Expression cloning of a protective *Leishmania* antigen. Science 1995;268(5210):563–6.
[15] Julia V, Rassoulzadegan M, Glaichenhaus N. Resistance to *Leishmania major* induced by tolerance to a single antigen. Science 1996;274(5286):421–3.
[16] Donnelly JJ, Ulmer JB, Shiver JW, Liu MA. DNA vaccines. Annu Rev Immunol 1997;15:617–48.
[17] Tighe H, Corr M, Roman M, Raz M. Gene vaccination: plasmid DNA is more than just a blueprint. Immunol Today 1998;19(2):89–97.
[18] Moss B. Genetically engineered poxviruses for recombinant gene expression. Proc Nat1 Acad Sci USA 1996;93:11341–8.
[19] Paoletti E. Applications of poxvirus vectors to vaccination: an update. Proc Natl Acad Sci USA 1996;93:11349–53.
[20] Rodríguez D, Rodríguez JR, Llorente M, et al. Use of DNA priming and vaccinia virus boosting to trigger an efficient immune response to HIV-1 gp120. Gene Therapy Mol Biol 1999;3:179–87.
[21] Zavala F, Rodrigues M, Rodríguez D, Rodríguez JR, Nussenzweig SR, Esteban M. A striking property of recombinant poxviruses: efficient inducers of in vivo expansion of primed $CD8^+$ T cells. Virology 2001;280:155–9.
[22] McMahon-Pratt D, Rodríguez D, Rodríguez JR, et al. Recombinant vaccinia viruses expressing GP46/M-2 protect against *Leishmania*infection. Infect Immunol 1993;61(8):3351–9.

[23] Gonzalo RM, del Real G, Rodriguez JR. A heterologous prime-boost regime using DNA and recombinant vaccinia virus expressing the *Leishmania infantum* P36/LACK antigen protects BALB/c mice from cutaneous leishmaniasis. Vaccine 2002;20(7/8): 1226–31.

[24] Schirmbeck R, Konig-Meredig SA, Riedl P, et al. Priming of immune responses to hepatitis B surface antigen with minimal DNA expression constructs modified with a nuclear localization signal peptide. J Mol Med 2001;79(5/6):343–50.

[25] Branden LJ, Mohamed AJ, Smith CI. A peptide nucleic acid-nuclear localization signal fusion that mediates nuclear transport of DNA. Nat Biotechnol 1999;17(8):784–7.

[26] Ludtke JJ, Zhang G, Sebestyen MG, Wolff JA. A nuclear localization signal can enhance both the nuclear transport and expression of 1 kb DNA. J Cell Sci 1999;112(Pt 12):2033–41.

[27] Zanta MA, Belguise-Valladier P, Behr JP. Gene delivery: a single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus. Proc Natl Acad Sci USA 1999;96(1):91–6.

[28] Chakrabarti S, Brechling K, Moss B. Vaccinia virus expression vector: coexpression of beta-galactosidase provides visual screening of recombinant virus plaques. Mol Cell Biol 1985;5:3403–9.

[29] Mougneau E, Altare F, Wakil AE. Expression cloning of a protective *Leishmania* antigen. Science 1995;268(5210):563–6.

[30] Carver TE, Lockey TD, Srinivas RV, Webster RG, Hurwitz JL. A novel vaccine regimen utilizing DNA, vaccinia virus and protein immunizations for HIV-1 envelope presentation. Vaccine 1999;17:1567–72.

[31] Hanke T, Blanchard TJ, Schneider J, et al. Enhancement of MHC class I-restricted peptide-specific T cell induction by a DNA prime/MVA boost vaccination regimen. Vaccine 1998;16:439–45.

[32] Hanke T, McMichael A. Pre-clinical development of a multi-CTL epitope-based DNA prime MVA boost vaccine for AIDS. Immunol Lett 1999;66:177–81.

[33] Hanke T, Samuel RV, Blanchard TJ, et al. Effective induction of simian immunodeficiency virus-specific cytotoxic T lymphocytes in macaques by using a multiepitopic gene and DNA prime-modified vaccinia virus Ankara boost vaccination regimen. J Virol 1999;73: 7524–32.

[34] Kent SJ, Zhao A, Best SJ, Chandler JD, Boyle DB, Ramshaw IA. Enhanced T cell immunogenicity and protective efficacy of a human immunodeficiency virus type 1 vaccine regimen consisting of consecutive priming with DNA and boosting with recombinant fowlpox virus. J Virol 1998;72:10180–8.

[35] Robinson HL, Montefiori D, Johnson P, et al. Neutralizing antibody-independent containment of immunodeficiency virus challenges by DNA priming and recombinant Pox virus booster immunizations. Nat Med 1999;5:526–34.

[36] Sedegah M, Jones TR, Kaur M, Hedstrom R, Hobart P, Tine JA, et al. Boosting with recombinant vaccinia increases immunogenicity and protective efficacy of malaria DNA vaccine. Proc Natl Acad Sci USA 1998;95:7648–53.

[37] Cressman DE, O'Connor WJ, Greer SF, Zhu XS, Ting JP. Mechanisms of nuclear import and export that control the subcellular localization of class II transactivator. J Immunol 2001;167(7):3626–34.

[38] Schneider J, Langermans JA, Gilbert SC, et al. A prime-boost immunisation regimen using DNA followed by recombinant-modified vaccinia virus Ankara induces strong cellular immune responses against the *Plasmodium falciparum* TRAP antigen in chimpanzees. Vaccine 2001;19(32):4595–602.

[39] Gherardi MM, Ramirez JC, Esteban M. Towards a new generation of vaccines: the cytokine IL-12 as an adjuvant to enhance cellular immune responses to pathogens during prime-booster vaccination regimens. Histol Histopathol 2001;16(2):655–67.

[40] Schneider J, Gilbert SC, et al. Induction of CD8+ T cells using heterologous prime-boost immunisation strategies. Immunol Rev 1999;170:29–38.

Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 91–96, January 1999
Biochemistry

Gene delivery: A single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus

MARIA ANTONIETTA ZANTA*, PASCALE BELGUISE-VALLADIER*, AND JEAN-PAUL BEHR†

Laboratoire de Chimie Génétique associé Centre National de la Recherche Scientifique/Université Louis Pasteur (Unité Mixte de Recherche 7514), Faculté de Pharmacie de Strasbourg, France

Communicated by Jean-Marie P. Lehn, Université Louis Pasteur, Strasbourg, France, November 3, 1998 (received for review September 9, 1998)

ABSTRACT **Translocation of exogenous DNA through the nuclear membrane is a major concern of gene delivery technologies. To take advantage of the cellular import machinery, we have synthesized a capped 3.3-kbp CMVLuciferase-NLS gene containing a single nuclear localization signal peptide (PKKKRKVEDPYC). Transfection of cells with the tagged gene remained effective down to nanogram amounts of DNA. Transfection enhancement (10- to 1,000-fold) as a result of the signal peptide was observed irrespective of the cationic vector or the cell type used. A lysine to threonine mutation of the third NLS amino acid completely abolished these remarkable features, suggesting importin-mediated translocation. Our hypothesis is that the 3-nm-wide DNA present in the cytoplasm is initially docked to and translocated through a nuclear pore by the nuclear import machinery. As DNA enters the nucleus, it is quickly condensed into a chromatin-like structure, which provides a mechanism for threading the remaining worm-like molecule through the pore. A single NLS signal is thus sufficient, whereas many signals on a gene would actually inhibit entry, the same DNA molecule being threaded through adjacent pores.**

The nuclear membrane of eukaryotic cells is freely permeable to solutes of up to about 9 nm (e.g., 40- 60-kDa proteins). Transport of larger molecules through nuclear pores is signal-mediated, involves shuttle molecules, and requires energy. The basic peptide derived from the simian virus 40 large tumor antigen (PKKKRKV) is a nuclear localization signal (NLS) that mediates binding of the karyophilic protein to importin α (1). Complex formation triggers binding to importin β and the ternary complex is then carried through the nuclear pore with the help of the GTPase Ran. Macromolecules and particles up to 25 nm have been shown to enter the nucleus in this way, although the translocation mechanism is still not fully understood (2–5).

Eukaryotic DNA viruses, which replicate in the nucleus, seem to be capable of diverting the cell's nuclear import machinery to their own benefit (6, 7). Recombinant viruses derived from these viruses are being used to carry therapeutic genes into humans. However, the host's immune response is currently a limitation to the clinical development of viral gene therapy. Nonviral alternatives using plasmid DNA and cationic carrier molecules suffer from a different drawback, the low efficacy of gene delivery. The main barrier to transgene expression *in vitro* is the nuclear membrane (8–10). Because breakdown of this membrane during cell division helps nuclear localization, this obstacle is probably still more a problem *in vivo*, where cells can be considered to be resting with respect to the lifetime of DNA.

Several attempts to improve entry of plasmid DNA into the nucleus have been published. These include use of electrostatic binding of DNA to cationic NLS-containing proteins (11, 12) or peptides (13) or lipids (14), as well as sequence-specific binding of DNA to karyophilic proteins (15–17). Yet such "piggyback" nuclear transport relies on the unpredictable stability of the complexes within the cytoplasm. Recent work by the group of Jon Wolff (18) with digitonin-permeabilized cells demonstrated nuclear accumulation of fluorescently labeled DNA that was randomly tagged with hundreds of NLS peptides; nuclei of intact cells did not take up the modified DNA. Our approach took advantage of a chemically controlled pathway to irreversibly link a single NLS-peptide to one end of a gene. Moreover, following Heisenberg's principle, we avoided taking a too close look at nuclear import using potentially disturbing techniques. The goal being gene delivery, the level of transgene expression after transfection was chosen as the most pertinent test of success.

EXPERIMENTAL PROCEDURES

Chemicals, Enzymes, and Oligonucleotides. 4-(*N*-Maleimidomethyl)cyclohexane-1-carboxylic acid *N*-hydroxysuccinimide ester (SMCC) was purchased from Sigma. The SAL 34-mer 5′-d(TCGATGTCCGCGTTGGCTTXTGCCAACGCGGACA) oligodeoxynucleotide was synthesized by Appligene by using an amino-modified deoxythymidine (X; amino-modified dT, Glen Research, Sterling, VA). The XMA 34-mer 5′-d(CCGGCTACCTTGCGAGCTTTTGCTCGCAAGGTAG) oligodeoxynucleotide, the NLS peptide NH_2-PKKKRKVEDPYC, and the mutated-NLS peptide NH_2-PKTKRKVEDPYC with C-terminal amidation were synthesized by Genosys (The Woodlands, TX). Hairpin structures composed of a loop of four thymines, a stem of 13 bp, and a sticky 5′ end were formed by boiling and subsequently cooling the XMA or the SAL oligonucleotides in ice. *XmaI*, *XmnI*, *SalI*, and *Bsp*HI restriction endonucleases, T4 polynucleotide kinase, T4 DNA ligase, and exonuclease III were purchased from New England Biolabs. Linear 22-kDa (ExGen500) and branched 25-kDa polyethylenimines (PEI) were purchased from Euromedex (Souffelweyersheim, France) and Fluka (Saint-Quentin Fallavier, France), respectively. Transfectam (dioctadecylamido glycylspermine) was synthesized as described (19).

Preparation of the Oligonucleotide-Peptide Conjugate. Two A_{260} units (6.46 nmol, assuming ε_{260} = 309,800 $M^{-1}cm^{-1}$) of the amino-modified SAL oligonucleotide in 20 μl of phosphate buffer [10 mM sodium phosphate (pH 7.5)] was mixed with a 40:1 molar excess of the bifunctional crosslinker SMCC in dimethylformamide (as a 30 mM stock solution) and incubated

Abbreviations: NLS, nuclear localization signal; SMCC, 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylic acid *N*-hydroxysuccinimide ester; PEI, polyethylenimine; N/P, nitrogen/phosphate.

*M.A.Z. and P.B.-V. contributed equally to this work.

†To whom reprint requests should be addressed at: Laboratoire de Chimie Génétique, Faculté de Pharmacie, BP 24, F-67401 Illkirch, France. e-mail: behr@aspirine.u-strasbg.fr.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

at room temperature for 2 h. Excess SMCC was removed with a Nick-Spin column (Amersham-Pharmacia) that had been equilibrated with PBS. The recovered oligonucleotide solution (100 μl) was immediately reacted with a 10-fold molar excess of NLS-peptide (or mutated-NLS-peptide) overnight at room temperature and then stored at −20°C. The oligonucleotide-peptide conjugate was purified by preparative PAGE. The coupling yield was 30%, based on quantification of the radiolabeled oligonucleotides after migration in a 20% denaturing polyacrylamide gel. To assess the peptide content, 1 pmol of radiolabeled oligonucleotide-NH_2 or oligonucleotide-NLS was incubated in 10 mM Tris·HCl, pH 8/1 mM EDTA in the presence of proteinase K (0.5 mg/ml) for 10 min at 37°C, followed by incubation for 10 min at 65°C and subsequent loading on a denaturing polyacrylamide gel. Visualization and quantification were performed with a PhosphorImager 425 (Molecular Dynamics).

Ligation of the CMVLuc Restriction Fragment to the Hairpin Oligonucleotides. pCMVLuc plasmid, encoding the *Photinus pyralis* luciferase under the control of the cytomegalovirus enhancer/promoter and followed by the simian virus 40 early polyadenylation signal (a gift from B. Demeneix, Museum National d'Histoire Naturelle, Paris) was propagated and purified as described (20). *Xmn*I/*Xma*I double digestion (10 U/μg of DNA) was performed at 37°C for 2 h, followed by enzyme heat inactivation (65°C for 20 min), before *Sal*I cleavage (37°C for 2 h; 10 U/μg of DNA). After removal of the endonucleases by DNA precipitation, *Bsp*HI digestion was performed for 2 h at 37°C (10 U/μg of DNA). Separation of the CMVLuc-containing DNA fragment (3,380 bp) from the shorter restriction fragments was performed by ultracentrifugation (30,000 rpm, SW41 rotor, 19 h at 25°C, Beckman Ultracentrifuge L8/55, France) in a 15–30% sucrose gradient.

The 5′ end of the oligonucleotide-peptide carrying the NLS or the mutated NLS peptide was radiolabeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (1 U/pmol of oligonucleotide at 37°C for 30 min). The 5′ end of the oligonucleotide-cap was phosphorylated similarly with ATP. Excess [γ-^{32}P]ATP and ATP were removed with Microspin G-25 columns (Amersham-Pharmacia). Before ligation, the hairpin form of the oligonucleotides presenting a sticky 5′ end was obtained by boiling and subsequently cooling the sample in ice. Ligation of the CMVLuc fragment with the oligonucleotide-cap and the oligonucleotide-peptide was performed overnight at 13°C with a 15-fold molar excess of each oligonucleotide and T4 DNA ligase (10,000 U/μg of DNA). The excess oligonucleotide was removed with a Microspin S-400 HR column (Amersham-Pharmacia). Quantification of the ligase reaction yield (80–90%) was performed by Cerenkov counting (TRI-CARB 2100 TR Liquid Scintillation Analyzer, Packard). Capping of the CMVLuc fragment was checked by digestion with exonuclease III (10 U/μg of DNA) at 37°C. Agarose gel electrophoresis showed the uncapped and hemicapped fragments to be totally digested, whereas the capped fragment remained undigested.

Cells and Cell Culture. NIH 3T3 murine fibroblasts were purchased from American Type Culture Collection and grown in DMEM (GIBCO/BRL). BNL CL.2 murine hepatocytes were provided by E. Wagner (Bender, Wien, Austria) and grown in DMEM, high glucose (4.5 g/liter). HeLa human cervix epitheloid carcinoma cells (provided by L. Monaco, Istituto San Raffaele, Milano, Italy) were grown in MEM with Earle's salt (PolyLabo, Strasbourg, France). Human monocytes were isolated from the blood of a healthy donor by elutriation/cytapheresis and differentiated into macrophages by 1-week culture in RPMI 1640 medium supplemented with 100 U/ml GM-CSF. Dorsal root ganglia neurons from new born rats were obtained and grown as described (21). Cell culture media were supplemented with 10% fetal calf serum (GIBCO/BRL), 2 mM L-glutamine, penicillin (100 units/ml), and streptomycin (100 μg/ml; GIBCO/BRL). Cells were maintained at 37°C in a 5% CO_2 humidified air atmosphere.

Cell Transfection. For each cell line used, 10,000 cells per well were seeded 24 h before transfection in 96-well tissue culture plates (Costar) to reach 60–70% confluence during transfection. Before transfection, cells were rinsed and 0.2 ml of fresh culture medium supplemented (transfection in the presence of serum) or not (transfection in the absence of serum) with 10% fetal calf serum was added to each well. The desired amount of DNA was diluted into 46 μl (final volume) of 0.15 M NaCl or 5% glucose. The desired quantity of ExGen500, 25-kDa PEI, or Transfectam (from a 1 mM aqueous amine nitrogen stock solution of PEI or a 2 mM ethanolic stock solution of Transfectam) was then added to the DNA-containing solutions, vortex-mixed gently, and centrifuged. After 10 min, volumes corresponding to 10, 20, or 200 ng of CMVLuc fragment (10 ng/μl of DNA) or to the gene-number-corrected amount of plasmid DNA were added to the cells. The cell culture dish was immediately centrifuged (Sigma product) for 5 min at 1500 rpm (280 × *g*) or 500 rpm for primary neurons. After 2–3 h, 20 μl of fetal calf serum was added to the serum-free wells. Cells were cultured for 24 h and tested for reporter gene expression. All experiments were done in duplicate.

Luciferase Assay. Luciferase gene expression was measured by a luminescence assay. The culture medium was discarded and cell lysate was harvested after incubation of cells for 30 min at room temperature in 100 μl of Lysis Reagent 1× (Promega). The lysate was vortex-mixed gently and centrifuged for 5 min at 14,000 rpm at 4°C (Sigma 3K10). Twenty microliters of supernatant was diluted into 100 μl of luciferase reaction buffer (Promega) and the luminescence was inte-

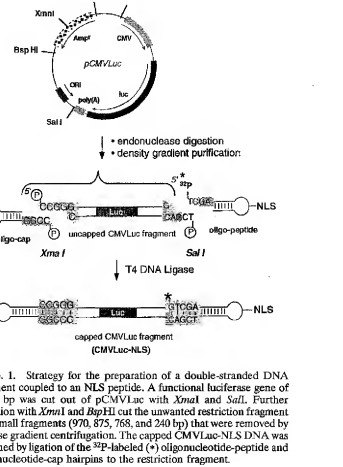

FIG. 1. Strategy for the preparation of a double-stranded DNA fragment coupled to an NLS peptide. A functional luciferase gene of 3,380 bp was cut out of pCMVLuc with *Xma*I and *Sal*I. Further digestion with *Xmn*I and *Bsp*HI cut the unwanted restriction fragment into small fragments (970, 875, 768, and 240 bp) that were removed by sucrose gradient centrifugation. The capped CMVLuc-NLS DNA was obtained by ligation of the ^{32}P-labeled (*) oligonucleotide-peptide and oligonucleotide-cap hairpins to the restriction fragment.

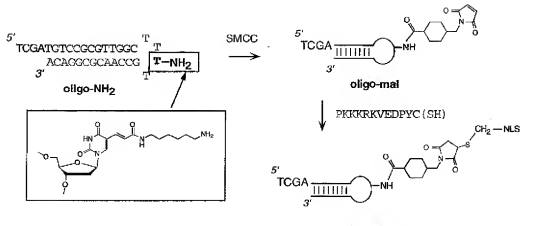

FIG. 2. Reaction scheme for the chemical coupling steps leading to the oligonucleotide-peptide conjugate (oligo-NLS). A hairpin oligonucleotide with a free allylamino group in the T_4 loop (oligo-NH_2) was reacted with the heterobifuncional crosslinker SMCC to give a thiol-reactive maleimide oligonucleotide (oligo-Mal), which was in turn reacted with the C-terminal cysteinamide residue of the NLS dodecapeptide.

grated over 10 sec (Mediators, Wien, Austria). Results were expressed as light units per mg of cell protein (BCA assay, Pierce).

RESULTS

Synthesis of a Capped Gene-Peptide Conjugate, Purification, and Proof of Structure. The construction was based on ligation of a pair of hairpin oligonucleotides to unique cohesive termini generated on the reporter gene (Fig. 1). Incorporation of ^{32}P into the modified oligonucleotide allowed us to follow the reaction kinetics, to purify the fragment of interest, and to verify its structure. The firefly luciferase reporter gene (Luc) flanked by the cytomegalovirus (CMV) enhancer/promoter sequence and the simian virus 40 polyadenylation signal was excised from the pCMVLuc plasmid (Fig. 1). Quadruple endonuclease digestion ensured straightforward large-scale separation of the 3,380-bp CMVLuc fragment from the remaining <1-kbp fragments by ultracentrifugation through a 15–30% sucrose gradient. T_4 loops are well suited for hairpins and the C-terminal thiol group of the PKKKRKVEDPYC peptide was conjugated to a thymine with a C(5) amino group (22) via an activated ester/maleimide bifunctional linker (SMCC) as detailed in Fig. 2. Preliminary experiments showed that the global reaction yield was highest after activation of oligonucleotide-NH_2 with SMCC for 2 h (significant parasitic maleimide hydrolysis may occur with time). The 5′ ^{32}P radiolabeling and polyacrylamide gel electrophoresis of oligonucleotides (Fig. 3*A*) showed the peptide conjugation reaction to be completed after 3 h, giving 30% oligonucleotide-NLS, based on the starting oligonucleotide. Proteinase K digestion converted oligonucleotide-NLS to a faster migrating compound, presumably the oligonucleotide conjugated to the C-terminal amino acid (oligonucleotide-Mal-Cys), thus establishing the chimeric nature of the conjugate (Fig. 3*B*). The oligonucleotide-NLS was purified by electrophoresis. The uncapped CMVLuc fragment was then simultaneously reacted with a 15-fold excess of oligonucleotide-cap and oligonucleotide-NLS by using T4 DNA ligase in previously optimized conditions. Ligation reaction yield (80–90%) was assessed by quantitative radioactivity counting and full capping of the CMVLuc fragment was checked by 3′ exonuclease digestion. CMVLuc-NLS was purified by gel permeation.

Transfection with Small Amounts of DNA. The reporter gene construct was obtained by chemical/enzymatic synthesis and purified by PAGE/gel permeation, instead of being amplified in bacteria. Only limited amounts of DNA could,

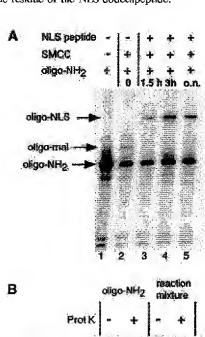

FIG. 3. (*A*) Synthesis of oligonucleotide-NLS. PAGE analysis of the formation of the oligonucleotide-peptide conjugate is shown. Aliquots of the reaction mixture were taken at different times and analyzed on a 20% denaturing gel after radiolabeling of the oligonucleotides. Lanes: 1, oligonucleotide-NH_2; 2, oligonucleotide-NH_2/SMCC reaction mixture after 2 h; 3–5, reaction mixture 1.5 h, 3 h or overnight after addition of the NLS-peptide to the oligonucleotide-NH_2/SMCC mixture, respectively. (*B*) The presumed oligonucleotide-NLS conjugate is a substrate of proteinase K. Radiolabeled oligonucleotide-NH_2 or the crude oligonucleotide-NLS reaction mixture were digested with proteinase K. Products were analyzed on a 20% denaturing gel and show total conversion of oligonucleotide-NLS into a faster migrating band, presumably oligo-Mal-Cys.

therefore, be obtained, and the transfection setup had to be miniaturized. A convenient 96-well microtiter plate assay developed by Felgner *et al.* (23) was chosen. Using several optimized cationic lipid formulations, these authors showed that transfection was best at 2–0.5 μg of plasmid and quickly fell off below 0.25 μg. We confirmed this result with pCMVLuc and two other cell types (Table 1, entry 4, and Fig. 4), thus putting our ultimate goal to obtaining an effective transfection with less than 200 ng of DNA. As a preliminary test, we looked at the relevance to cap the reporter gene at both ends, thus avoiding free DNA ends that are prone to exonuclease-mediated DNA degradation and to eliciting a DNA repair response. To this end, the uncapped, hemicapped, and fully capped CMVLuc genes (200 ng) were transfected into hepatocytes. Luciferase activities (Table 1) showed both un- and hemicapped molecules to give 25-fold less gene expression than the capped one thus justifying the choice of our chemical strategy.

The NLS Peptide Allows Effective Transfection with Minute Quantities of DNA. 3T3 cells were transfected with decreasing amounts of DNA—the NLS-bearing capped gene (CMVLuc-NLS), the capped gene (CMVLuc), and the corresponding mass-corrected amount of plasmid DNA (pCMVLuc). Although efficacy very much decreased for "20-ng" plasmid DNA (in fact 38 ng), as expected (see above), transfection levels remained remarkably high and constant over the range of 200–10 ng of DNA with CMVLuc-NLS (Fig. 4). Comparison with the capped gene lacking the NLS peptide showed 100- to 1,000-fold more expression when the NLS peptide was present on the gene. Similar conclusions were obtained, by using a cationic lipid (Transfectam, Fig. 4 *Left*) or a cationic polymer (branched PEI, Fig. 4 *Right*). Further data (not shown) confirmed that other cationic carrier molecules behaved similarly but also that efficacy dropped 50-fold within the range of 10–1 ng of CMVLuc-NLS.

Improved Transfection Involves the Cellular Nuclear Import Machinery. Comparative experiments between CMVLuc and CMVLuc-NLS, although interesting, do not really allow one put forward any hypothesis concerning the mechanism. Nuclear import of DNA and oligonucleotides has mainly been followed with fluorescent tags and fluorescence *in situ* hybridization, in microinjected or digitonin-permeabilized cells. Proof of active import relied on inhibition experiments with energy-decoupling molecules, nuclear pore-binding agglutinins, or mutations in the signal peptide sequence. Gene expression is an unambiguous proof of nuclear localization and transfection can be regarded as a straightforward way of introducing DNA into intact cells. One lysine to threonine mutation abolishes nuclear import (24). We therefore synthesized a capped gene (CMVLuc-mNLS) with a mutated PKTKRKVEDPYC sequence. Comparative transfection experiments with HeLa cells are shown in Fig. 5. Irrespective of the amount of DNA used, the lysine to threonine mutation brought transfection down to the level obtained with the capped gene having no peptide at all, confirming the functionality of the NLS–importin α interaction.

Table 1. Free DNA ends decrease efficacy

DNA form	Symbol	Transfection efficiency, RLU/mg of protein
Uncapped CMVLuc		$1 \pm 0.4 \times 10^5$
Hemi-capped CMVLuc		$9 \pm 2 \times 10^6$
CMVLuc		$2.5 \pm 0.6 \times 10^6$
Circular pCMVLuc plasmid		$2.5 \pm 0.2 \times 10^5$ $3 \pm 1 \times 10^{4*}$

BNL CL.2 hepatocytes were transfected with 200 ng of DNA per well. Complexes were formed with 25-kDa PEI (N/P = 10) in 150 mM NaCl and in the absence of serum. RLU, relative light units.
*Transfection with 20 ng of plasmid.

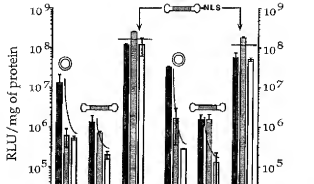

FIG. 4. NLS peptide promotes high and sustained transfection levels down to 10 ng DNA. 3T3 cells were transfected with decreasing amounts of DNA complexed in 150 mM NaCl to a cationic lipid (Transfectam, vector nitrogens over DNA phosphate, N/P = 6) or to a cationic polymer (25-kDa PEI, N/P = 10) in the absence of serum.

NLS Peptide-Mediated Transfection Enhancement Is a General Phenomenon. A series of comparative experiments was undertaken with various cell types. Transfection enhancement due to the presence of the NLS peptide was always observed (Table 2). However, enhancement factors were spread widely with cell type (10- to 1,000-fold), with no obvious relation with tissue origin nor cell type (primary vs. transformed cells). Nondividing primary cells such as human monocyte-derived macrophages and rat dorsal root ganglia neurons showed less impressive enhancement than 3T3 and HeLa cells. However, similar values were also seen for the easily transfected and fast dividing rat hepatocyte-derived cell line BNL CL.2.

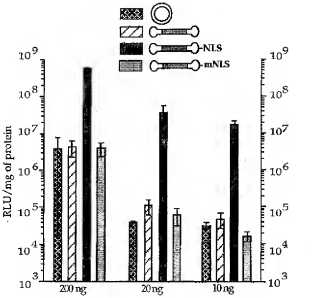

FIG. 5. Sustained luciferase expression levels are due to the nuclear localization peptide. HeLa cells were transfected with various amounts of DNA complexed to ExGen500 PEI (N/P = 5 in a 5% glucose solution) in the presence of 10% serum.

Table 2. Average enhancement of transfection with CMVLuc-NLS

Cells	Enhancement factor*
BNL CL.2	10
3T3	400†–1000
HeLa	200‡–300
Murine DRG neurons	30
Human macrophages	10

Values refer to transfection with 10–20 ng of DNA per well, in the absence of serum. DRG, dorsal root ganglion.
*Estimated spread is ±30%.
†In the presence of 10% serum during transfection.
‡In 24-well plates using 200 ng of DNA in the presence of 10% serum.

The general experimental setup for transfection included 96-well microtiter plates and no serum for 2 h after addition of the DNA–vector complexes to the cells. Several experiments were also performed on a larger scale (24-well plates) or in the presence of 10% serum during transfection. The results obtained for HeLa and 3T3 cells (Table 2) confirmed the general conclusions derived with the 96-well setup.

Time Course of the Transgene Expression. Active transport of the reporter gene via the nuclear import machinery could result in faster appearing expression. More interesting, a lower intracellular barrier to gene delivery means less DNA–vector complexes required in the cytoplasm and hence lower toxicity and more sustained expression. The kinetics of gene expression was followed in detail up to 24 h (Fig. 6) and then daily up to 3 days. Again, transfection with CMVLuc-NLS was much more effective than with CMVLuc-mNLS. Remarkably (and reproducibly), NLS-driven transfection reached its plateau after 12 h, whereas transfection levels obtained with the plasmid or the mutated NLS sequence were still increasing significantly up to 24 h. Previous experiments (25) showed both transfection and cytoplasmic injection of DNA–PEI complexes to have superimposable kinetics of transgene expression, suggesting that the slow step was intracellular trafficking/nuclear entry rather than cell entry. Using the nuclear import machinery seems to speed up the rate-limiting step. For longer time periods, expression stayed broadly constant with a low standard deviation. However, the rather large variability observed upon repeating this experiment prevented us from drawing any conclusion (the inconsistent results observed at $t > 1$ day may be due to the fact that cells reached confluency at day 1).

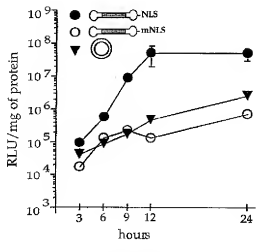

FIG. 6. Reporter protein activity appears faster with CMVLuc-NLS. HeLa cells were transfected with 10 ng of DNA complexed to ExGen 500 PEI (N/P = 5 in 5% glucose solution) in the absence of serum. Cells were lysed and luciferase activity was measured at the indicated time after transfection. The absence of error bars indicates that the SEM is smaller than the label on the graph.

DISCUSSION

The goal of this work was to improve nonviral plasmid-mediated gene delivery. When fully counterion-condensed, a single plasmid molecule collapses into a sphere of about 25 nm (26) that is already at the size-exclusion limit for signal-mediated nuclear import (4, 27). Plasmid condensation with cationic lipids or polymers generally leads to even larger multimolecular aggregates that reach the cytoplasm after binding to cell-surface anionic proteoglycans (9, 28) and eventual escape from the formed vacuoles (8, 9, 30). Because the particles are too large to cross an intact nuclear membrane, transfection of resting cells can reasonably only be accounted for by uncomplexed DNA that has been released by exchange with phosphatidylserine (31) or heparan sulfate (9) present in the vacuolar membrane. Any molecular information (e.g., NLS) borne by the cationic vector will thus be lost before reaching the nuclear membrane, hence its ineffectiveness.

In sharp contrast to a condensed DNA molecule, a free hydrated DNA double helix (3 nm) is thin enough to enter the nuclear pore by diffusion, i.e., with no energy or signal requirement. Restricted intracellular motion of DNA [the cytoplasm is equivalent to 13% dextran (32)], and a rather low nuclear pore coverage [<10% of the membrane surface (33)], however, give this event a low probability to occur: the nuclear membrane has been recognized as a major barrier to gene delivery (8, 9). Quantitative cytoplasmic microinjection of DNA (25, 34, 35) indeed led to less than 0.1% of the DNA being expressed. The probability for DNA to find and subsequently enter a nuclear pore can be increased by a bound karyophilic signal peptide able to dock DNA to the nuclear pore filaments and help an initial part of the molecule to cross the membrane. The DNA–karyophilic signal link should, however, be stable, hence the chemical strategy we chose to use in this study.

The major achievement of our approach is that the amount of DNA required to effectively transfect cells *in vitro* has now shifted from the microgram to the nanogram range (see also ref. 17). Although the improvement (one to three orders of magnitude) varies with cell type (showing that the nuclear membrane is not the only possible barrier), even the modest

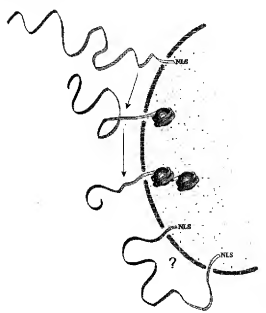

FIG. 7. Hypothetical scheme of foreign DNA entering the nucleus. The vectorial movement is driven by DNA binding to chromatin-forming basic proteins. A single NLS peptide helps initial threading, whereas too many signals may inhibit full translocation.

enhancements observed for primary cell cultures look interesting (1,000%) when not presented on a logarithmic scale. Is this exciting result really a consequence of the involvement of the nuclear import machinery? We choose to tackle the academic aspect of this work by using the lysine to threonine mutation, which indeed abolished the beneficial effect brought about by the DNA-bound NLS sequence.

Taking our hypothesis further, the last unanswered question deals with threading of the rest of the plasmid molecule (Fig. 7). Random-walk diffusion of a micrometer-long molecule would be inefficient and slow. Fortunately, naked DNA will not remain free in the nucleus: histones (and eventually basic nuclear matrix proteins) indeed quickly assemble transfected DNA into chromatin-like structures (36, 37), thus providing a mechanism for pulling and condensing the filamentous molecule into the nucleus. Following this line of thinking, many NLS signals distributed along the DNA may actually inhibit nuclear entry if the nucleic acid is longer than the distance separating adjacent pores (Fig. 7). A straightforward calculation that uses the pore density in HeLa cells (33) shows this to become probable above 1 kbp.

The present work is a proof of principle rather than a new gene delivery technology. However, it will hopefully provide the impetus for searching more versatile solutions to the link between a plasmid and a NLS peptide, such as triple-helix forming (38, 39) or strand-invading (40, 41) oligonucleotide-NLS conjugates.

We thank Dr. Olivier Feugeas and Dr. Anne Feltz for providing the monocytes and neurons, respectively. We are grateful to Dr. Robert P. P. Fuchs and Dr. M. Bichara for providing technical facilities and for helpful discussions. This work was supported by grants from the Association Française contre les Myopathies, the Association pour la Recherche contre le Cancer, and the Ligue contre le Cancer. M.A.Z. received a fellowship from Transgene SA.

1. Adam, S. A. & Gerace, L. (1991) *Cell* **66**, 837–847.
2. Panté, N. & Aebi, U. (1995) *J. Cell Sci. Suppl.* **19**, 1–11.
3. Nigg, E. A. (1997) *Nature (London)* **386**, 779–787.
4. Ohno, M., Fornerod, M. & Mattaj, I. W. (1998) *Cell* **92**, 327–336.
5. Melchior, F. & Gerace, L. (1998) *Trends Cell. Biol.* **8**, 175–179.
6. Greber, U. F., Suomalainen, M., Stidwill, R. P., Boucke, K., Ebersold, M. W. & Helenius, A. (1997) *EMBO J.* **16**, 5998–6007.
7. Greber, U. F. (1998) in *Self-Assembling Complexes for Gene Delivery: From Laboratory to Clinical Trial*, eds. Kabanov, A. V., Felgner, P. L. & Seymour, L. W. (Wiley, Chichester), pp. 89–114.
8. Zabner, J., Fasbender, A. J., Moninger, T., Poellinger, K. A. & Welsh, M. J. (1995) *J. Biol. Chem.* **270**, 18997–19007.
9. Labat-Moleur, F., Steffan, A. M., Brisson, C., Perron, H., Feugeas, O., Furstenberger, P., Oberling, F., Brambilla, E. & Behr, J. P. (1996) *Gene Therapy* **3**, 1010–1017.
10. Hagstrom, J. E., Ludtke, J. J., Bassik, M. C., Sebestyén, M. G., Adam, S. A. & Wolff, J. A. (1997) *J. Cell Sci.* **110**, 2323–2331.
11. Kaneda, Y., Iwai, K. & Uchida, T. (1989) *Science* **243**, 375–378.
12. Fritz, J. D., Herweijer, H., Zhang, G. F. & Wolff, J. A. (1996) *Hum. Gene Ther.* **7**, 1395–1404.
13. Coffas, P., Husebye, H. & Aleström, P. (1996) *Transgenic Res.* **5**, 541–548.
14. Remy, J. S., Kichler, A., Mordvinov, V., Schuber, F. & Behr, J. P. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 1744–1748.
15. Fominaya, J. & Wels, W. (1996) *J. Biol. Chem.* **271**, 10560–10568.
16. Dean, D. A. (1997) *Exp. Cell Res.* **230**, 293–302.
17. Langle-Rouault, F., Patzel, V., Benavente, A., Taillez, M., Silvestre, N., Bompard, A., Sczakiel, G., Jacobs, E. & Rittner, K. (1998) *J. Virol.* **72**, 6181–6185.
18. Sebestyén, M. G., Ludtke, J. J., Bassik, M. C., Zhang, G., Budker, V., Lukhtanov, E. A., Hagstrom, J. E. & Wolff, J. A. (1998) *Nat. Biotech.* **16**, 80–85.
19. Behr, J. P., Demeneix, B., Loeffler, J. P. & Perez-Mutul, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6982–6986.
20. Zanta, M. A., Boussif, O., Adib, A. & Behr, J. P. (1997) *Bioconjugate Chem.* **8**, 839–844.
21. Lambert, R. C., Maulet, Y., Dupont, J. L., Mykita, S., Craig, P., Volsen, S. & Feltz, A. (1996) *Mol. Cell. Neurosci.* **7**, 239–246.
22. Seibel, P., Trappe, J., Villani, G., Klopstock, T., Papa, S. & Reichmann, H. (1995) *Nucleic Acids Res.* **23**, 10–17.
23. Felgner, J. H., Kumar, R., Sridhar, C. N., Wheeler, C. J., Tsai, Y. J., Border, R., Ramsey, P., Martin, M. & Felgner, P. L. (1994) *J. Biol. Chem.* **269**, 2550–2561.
24. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. (1984) *Cell* **39**, 499–509.
25. Pollard, H., Remy, J. S., Loussouarn, G., Demolombe, S., Behr, J. P. & Escande, D. (1998) *J. Biol. Chem.* **273**, 7507–7511.
26. Blessing, T., Remy, J. S. & Behr, J. P. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 1427–1431.
27. Görlich, D. & Mattaj, I. W. (1996) *Science* **271**, 1513–1518.
28. Mislick, K. A. & Baldeschwieler, J. D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12349–12354.
29. Plank, C., Oberhauser, B., Mechtler, K., Koch, C. & Wagner, E. (1994) *J. Biol. Chem.* **269**, 12918–12924.
30. Boussif, O., Lezoualch, F., Zanta, M. A., Mergny, M. D., Scherman, D., Demeneix, B. & Behr, J. P. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7297–7301.
31. Xu, Y. H. & Szoka, F. C. (1996) *Biochemistry* **35**, 5616–5623.
32. Luby-Phelps, K. (1994) *Curr. Opin. Cell Biol.* **6**, 3–9.
33. Maul, G. G. & Deaven, L. (1977) *J. Cell Biol.* **73**, 748–760.
34. Mirzayans, R., Aubin, R. A. & Paterson, M. C. (1992) *Mutat. Res.* **281**, 115–122.
35. Dowty, M. E., Williams, P., Zhang, G., Hagstrom, J. E. & Wolff, J. A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 4572–4576.
36. Cereghini, S. & Yaniv, M. (1984) *EMBO J.* **3**, 1243–1253.
37. Jeong, S. & Stein, A. (1994) *Nucleic Acids Res.* **22**, 370–375.
38. Moser, H. E. & Dervan, P. B. (1987) *Science* **238**, 645.
39. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J., Thuong, N. T., Lhomme, J. & Hélène, C. (1987) *Nucleic Acids Res.* **15**, 7749.
40. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. (1991) *Science* **254**, 1497–1500.
41. Schmid, N. & Behr, J. P. (1995) *Tetrahedron Lett.* **36**, 1447–1450.

Cell, Vol. 39, 499–509, December 1984 (Part 2), Copyright © 1984 by MIT

0092-8674/84/130499-11 $02.00/0

A Short Amino Acid Sequence Able to Specify Nuclear Location

Daniel Kalderon, Bruce L. Roberts, William D. Richardson, and Alan E. Smith
Biochemistry Division
National Institute for Medical Research
Mill Hill
London NW7 1AA, England

Summary

A short sequence of amino acids including Lys-128 is required for the normal nuclear accumulation of wild-type and deleted forms of SV40 large T antigen. A cytoplasmic large T mutant that lacks sequences from around Lys-128 localizes to the nucleus if the missing sequence is attached to its amino terminus. The implication that the sequence element around Lys-128 acts as an autonomous signal capable of specifying nuclear location was tested directly by transferring it to the amino termini of β-galactosidase and of pyruvate kinase, normally a cytoplasmic protein. Sequences that included the putative signal induced each of the fusion proteins to accumulate completely in the nucleus but had no discernible effect when Lys-128 was replaced by Thr. By reducing the size of the transposed sequence we conclude that Pro-Lys-Lys-Lys-Arg-Lys-Val can act as a nuclear location signal. The sequence may represent a prototype of similar sequences in other nuclear proteins.

Introduction

The nuclear and cytoplasmic compartments of eucaryotic cells contain largely distinct sets of proteins (Bonner, 1975; De Robertis et al., 1978). Studies of the fate of endogenously synthesized and exogenously introduced proteins suggest that all nuclear proteins are synthesized in the cytoplasm, from which the mature polypeptides migrate rapidly to the nucleus (Gurdon, 1970; Craig and Perry, 1971; Wu and Warner, 1971; Bonner, 1975; Feldherr, 1975; De Robertis et al., 1978; Yamaizumi et al., 1978). Thus the steady-state segregation of proteins between nucleus and cytoplasm appears to be governed by an intrinsic property of mature polypeptides. How the structures of proteins specify their distribution between the two compartments is not known.

Passage of proteins across the membranes of the endoplasmic reticulum, of mitochondria, or of chloroplasts is dictated by signal sequences of 20–40 contiguous amino acids (Blobel, 1980; Watson, 1984). Specific signals are not, however, necessarily required for passage across the nuclear envelope. For instance, many large polysaccharides and several proteins that are normally secreted or membrane-bound have been observed to enter the nucleus following introduction into the cytoplasm, and to enter at a rate apparently governed only by their size (Paine and Feldherr, 1972). Thus passage of proteins below M, 15,000 is not detectably retarded, a globular protein of M, 67,000 (BSA) enters nuclei only very slowly, and a protein of M, 450,000 (ferritin) is excluded (Paine et al., 1975). From such studies, the nuclear envelope has been proposed to behave as an array of cylindrical pores of internal diameter 7–11 nm (Paine et al., 1975; Peters, 1984). This model of the envelope corresponds reasonably well with the dimensions of discrete structures called nuclear pores, which have been purified from the nuclear envelope and analyzed by electron microscopy (Harris, 1981; Unwin and Milligan, 1982).

It appears therefore that in vivo many small proteins may diffuse freely between nucleus and cytoplasm and accumulate in either compartment due to selective retention or differential degradation. However, some means of entry to the nucleus other than diffusion must exist for large proteins, notwithstanding the possibility that some such proteins might invade the nuclear area during mitosis, at which time the nuclear envelope breaks down transiently in most eucaryotes. It seems likely, therefore, that access of the majority of large proteins to the nucleus requires that they participate in a specific interaction with a cellular component involved in transport of proteins to the nucleus. The specificity of this interaction could then account for the selective distribution of such proteins between nucleus and cytoplasm.

To date, two approaches have been taken to search for nuclear location signals in proteins. One approach assesses which parts of a protein are required for nuclear localization and is best exemplified by studies on the subcellular location of partial proteolytic digestion products of the nuclear protein nucleoplasmin (Dingwall et al., 1982). The other approach seeks to determine which parts of a nuclear protein are sufficient to promote nuclear accumulation. Thus, Hall and co-workers constructed chimeric fusion proteins to define which amino acid sequences of the yeast *Mat* α2 gene product sufficed to target E. coli β-galactosidase to the nucleus in yeast (Hall et al., 1984). Here we have combined these approaches to demonstrate that a very short sequence of amino acids derived from SV40 large T is required for its normal nuclear localization and promotes the nuclear accumulation of an otherwise cytoplasmic protein, pyruvate kinase and also of a bacterial protein, β-galactosidase.

Results

Sequences Required for Nuclear Location of Large T

It has previously been shown that alteration of Lys-128 of SV40 large T to Thr or Asn (Kalderon et al., 1984; Lanford and Butel, 1984) prevents the normal nuclear accumulation of large T in established rodent and simian cells without grossly altering the stability of the protein or its associated biochemical activities (Lanford and Butel, 1980; Paucha et al., submitted). Alteration of each of the basic residues adjacent to Lys-128 impaired but did not prevent nuclear accumulation. Other mutants with deletions bordering the

basic tract of amino acids around Lys-128 were wholly or predominantly nuclear, leading to the suggestion that a very short sequence of amino acids including Lys-128 may act as an independent element within large T and be responsible for bringing about its normal nuclear localization (Kalderon et al., 1984).

Here we have tested this hypothesis by constructing a series of mutants encoding proteins that lack sequences from the normal amino terminus or carboxy terminus, or from the regions on either side of Lys-128 in the primary structure of large T (Figure 1 and section 1 of Experimental Procedures).

The subcellular localization of variant large T proteins was examined by immunofluorescent staining 18 hr after needle microinjection of plasmid DNA into Vero cells and also, for those mutants still able to transform Rat-1 cells morphologically, in cell lines derived from isolated transformed dense foci. The results showed that nuclear accumulation was not affected by deletion of amino acid residues 1 to 82 (and perhaps 1 to 108) or 136 to 250 and was only slightly impaired by deletion of residues 98 to 126 (Figure 2). Furthermore, proteins truncated from the carboxy terminus as far as residue 178 were also located wholly in the nucleus. This set of mutants therefore collectively deletes most of the primary sequence of large T while sparing the element around Lys-128 without visibly impairing the ability of the mutant proteins to accumulate in nuclei.

However, as mechanisms of nuclear accumulation may differ among proteins (De Robertis, 1983) it is possible that wild-type large T and mutant forms of the protein distribute between nucleus and cytoplasm by different means. The nuclear accumulation of wild-type large T appears to depend on a mechanism involving Lys-128 (Kalderon et al., 1984). To test whether the observed subcellular location of each of the large T variants was also dependent on this mechanism, the product of equivalent proteins that included Thr rather than Lys at position 128 was examined (sections 1 and 2 of Experimental Procedures). Several of these Thr-128-containing mutants gave rise to proteins that were detected only in the cytoplasm (Figure 2), implying that the Lys-128 versions of the proteins depend on the same mechanism as wild-type large T for nuclear location.

By contrast, the 272 amino acid long truncated version of d10 large T (encoding Thr at 128) accumulated in the nucleus to almost the same extent as its wild-type counterpart. Indeed, every other truncated version of d10 large T that we tested (of lengths 361, 437, ca. 540, and ca. 620 amino acids), although predominantly cytoplasmic, was present to some extent in the nucleus. It is therefore clear from the difference in the distribution of d10 and wild-type versions of the truncated proteins that the Lys-128-dependent mechanism is operative in each case but it also appears that an additional means of nuclear accumulation is available to these proteins. It may be that the reduced size of the variants allows them to diffuse across the nuclear envelope.

Thus a Lys-128-dependent mechanism contributed

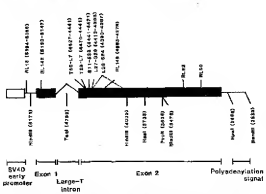

Figure 1. Linker Insertion/Deletion Mutations within the Large T Gene

The distribution of native restriction enzyme sites in the SV40 large T gene and the positions of insertion of a 10 bp Eco RI linker (CCGAATTCGG) in various derivatives of the wild-type gene are shown. The insertion sites are in each case denoted as the first SV40 nucleotide preceding or following the linker, in the numbering system of Tooze (1981). These enzyme sites were used as described in section 1 of Experimental Procedures to construct plasmids encoding deleted, duplicated, or truncated versions of large T. Note that RL52 and 60 are not precisely mapped.

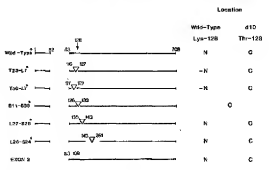

Figure 2. Structure and Subcellular Location of Large T Deletion Mutants

The derivation of all plasmids whose coding potential and properties are displayed above is described in sections 1 and 2 of Experimental Procedures. In each case only the native large T amino acid sequence encoded by the mutant plasmids is shown. Thus the linker insertion/deletion mutants also include three or four amino acids encoded by the Eco RI linker (CGGAATTCGG) inserted at the sites of the deletions shown. TL Pvu II also encodes one and TL 148 two additional amino acids before termination of translation. The location of all mutant proteins was ascertained by indirect immunofluorescence as described previously (Kalderon et al., 1984) using as first antibody either a monoclonal that recognizes amino-terminal (PAb419) or carboxy-terminal (PAb429) regions of large T (Lane et al., 1981) 18 hr after needle microinjection of DNA into Vero cells. For those plasmids marked with an asterisk, the distribution of mutant protein was also examined by indirect immunofluorescence using hamster anti-SV40 tumor serum as first antibody in Rat-1 cells transformed by the mutant proteins. The symbols N and C refer to complete nuclear or cytoplasmic localization; ~N indicates that the majority of protein is seen in the nucleus with only a trace in the cytoplasm; N > C and C > N indicate the relative intensities of fluorescence in nuclear and cytoplasmic compartments.

either wholly or in part to the nuclear accumulation of all the deleted and truncated forms of large T examined. Because it operates in the context of so many variant structures, this result suggests that the sequence around Lys-128 does not rely on an interaction with other parts of the molecule to ensure the nuclear location of large T.

Nuclear Accumulation of Large T Is Positively Promoted by Sequences Including Lys-128

It has been assumed above that Lys-128 forms part of a positive signal that promotes nuclear localization, rather than Thr-128 inducing cytoplasmic retention of large T. The latter possibility seemed unlikely because a number of deletion mutants that remove Lys-128 and adjacent sequences also share the cytoplasmic phenotype (Kalderon et al., 1984). Nevertheless, this possibility was examined directly by constructing large T variants that contained two successive copies of a short sequence (amino acids 127 to 135) around residue 128 with either Thr or Lys at this position (Figure 3). In every case where a wild-type sequence (including Lys at 128) was present, the variant protein localized exclusively to the nucleus, irrespective of whether this sequence was preceded or followed by sequences derived from d10 large T (Figures 3, 4a, 4c, 4d). The same result was observed when the size of the duplication was increased to 21 amino acids (127 to 147), although in this case only one of the two orientations of wild-type relative to mutant sequence was examined (data not shown). Thus the presence of Thr-128 in this position relative to the normal amino or carboxy terminus of large T does not necessarily induce cytoplasmic retention, whereas nuclear accumulation is promoted in every case when Lys occupies position 128.

The unimpaired nuclear accumulation of molecules with small duplications (Figure 4a, 4c, 4d) provides further evidence, similar to that derived from analysis of deletion mutants, that the function of the region around Lys-128 is not critically dependent on its position relative to the rest of the molecule.

Relocation of Putative Nuclear Location Signal to the Amino Terminus of Large T

The insertion/deletion mutant S11-S33 lacks native large T amino acid residues between 127 and 132 and fails to localize to the nucleus (Kalderon et al., 1984). We sought to restore a nuclear location to this protein by introducing at the amino terminus a short sequence of amino acids from around Lys-128 in wild-type large T (Figure 3 and section 4 of Experimental Procedures). Again, an analogous construction was made in which Lys-128 was replaced by a Thr residue. The subcellular distribution of these proteins was examined following microinjection of plasmid DNA into Vero cells by immunofluorescent staining using a monoclonal antibody directed against the amino terminus of large T PAb419 (Harlow et al., 1981). Virtually all of the chimeric protein with an amino-terminal basic tract including Lys-128 was seen in the nucleus, whereas the Thr-containing large T variant was entirely cytoplasmic

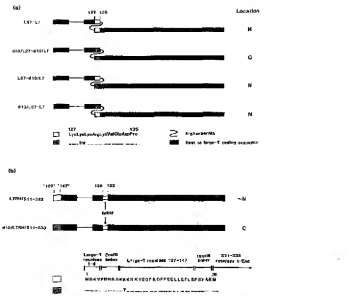

Figure 3. Structure and Subcellular Location of Large T Variants Containing Two Copies or One Relocated Copy of a Short Sequence around Residue 128

The construction of plasmids encoding large T variants in which sequences around Lys-128 were (a) duplicated or (b) transposed from their normal location to the amino terminus is described in sections 1, 3, and 4 of Experimental Procedures. The amino acid sequence following Met-36 of L7Pi-i/S11-S33 is identical with that following Met-1 of the linker insertion/deletion mutant, S11-S33 (which lacks residues 127 to 132). The location of variant proteins was assessed following microinjection of plasmid DNA into Vero cells by indirect immunofluorescence microscopy (see Figure 4).

(Figures 4e and 4f). Thus a short sequence around Lys-128 is able to confer a nuclear location to large T even when moved by over a hundred amino acids within the primary structure.

The tolerance of this sequence to transposition suggests that it constitutes an independent structural element but does not rule out the alternative possibility that its function depends on an interaction with other regions of the large T molecule. The hypothesis that the sequence acts as an autonomous nuclear localization signal was therefore tested more rigorously by ascertaining whether it is sufficient to promote the nuclear accumulation of proteins that are normally found in the cytoplasm. Plasmids encoding chimeric proteins were therefore constructed in which the putative nuclear location signal was close to a terminus of the test protein where, it could be argued, there is most chance of maintaining a structure that is independent of the rest of the molecule.

Localization of Chimeric β-Galactosidase Proteins

β-galactosidase, encoded by the *lac Z* gene of *E. coli*, was initially selected as a suitable test protein for several reasons. As a bacterial protein it would not be expected to include specific signals that facilitate transport across the nuclear envelope, although it may have an affinity for

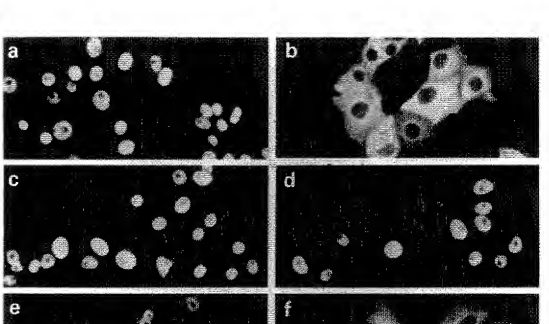

Figure 4. Subcellular Locations of Large T Variants Containing Two Copies, or One Relocated Copy of a Short Sequence around Residue 126

Vero cells were microinjected with plasmid DNAs encoding the variant large T molecules shown in Figure 3. Following microinjection, cells were fixed and stained using monoclonals PAb423 (a to d) or PAb419 (e and f) (Harlow et al., 1981) as first antibody. The second antibody was rhodamine-labeled rabbit anti-mouse IgG (Fab, fragment). Micrographs a to d represent large T variants containing two consecutive copies of the amino acid sequence from 127 to 135 with either Lys or Thr at position 128, as follows: (a) Lys-128 in both copies, L27-L7; (b) Thr-128 in both copies, d10/L27-d10/L7; (c) Lys-128 followed by Thr-128, L27-d10/L7; (d) Thr-128 followed by Lys-128, d10/L27-L7. Micrographs (e) and (f) show the localization of large T derivatives in which amino acids 127 to 147 including either Lys or Thr at position 128 have been added to the amino terminus of the cytoplasmic mutant 511-533 as follows: (e) Lys at position 128, L7P4/511-533; (f) Thr at position 128, d10/L7P4/511-533.

nuclear components common to procaryotic and eucaryotic cells. The large size of the monomeric protein (M, 116,000; Fowler and Zabin, 1977) and its propensity to oligomerize (Wallenfels and Weil, 1972) might be expected to preclude rapid entry to the nucleus in the absence of a specific transport mechanism. Furthermore, of the β-galactosidase fusion proteins that have previously been constructed, many retain enzyme activity (Casadaban et al., 1983), indicating that additional amino acid sequences can be accommodated at either the amino or carboxy terminus of the protein without gross effect on its native structure. Finally, the same approach has been used previously to identify sequences that localize β-galactosidase to the nucleus in yeast (Hall et al., 1984).

Initially, four types of plasmid encoding β-galactosidase fusion proteins were made (section 5 in Experimental Procedures). All contained the same promoter, translation initiation, and polyadenylation signals (all derived from the SV40 early region) and the same β-galactosidase coding sequence (Figure 5). They differed only in the number of SV40 large T amino acids that preceded the fusion junction with β-galactosidase. The subcellular location of each of the fusion proteins was examined, following microinjection of the appropriate plasmid DNA into Vero cells, by indirect immunofluorescence using a cocktail of four monoclonal antibodies raised against E. coli β-galactosidase (a gift from Dr. J. Partaledis).

Surprisingly, the hybrid protein that included only four amino acids from the amino terminus of SV40 large T (SRL30-BGAL, Figure 5) was seen at approximately equal intensities in both nuclear and cytoplasmic compartments. We obtained the same result with a hybrid protein, pCH110 (Hall et al., 1983) that included 40 amino acids encoded by the E. coli *gpt* gene and 28 by the *trp S* gene at the amino terminus of β-galactosidase. Similar results have been obtained by others for a series of amino- and carboxy-terminal β-galactosidase fusion proteins (D. P. Lane, personal communication). Thus, despite its large size and

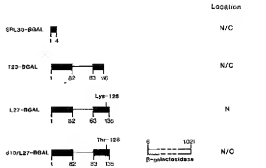

Figure 5. Structure and Subcellular Location of β-Galactosidase Fusion Proteins

Plasmids encoding a variable number of amino acid residues derived from the amino terminus of large T fused to β-galactosidase (see section 5 of Experimental Procedures for plasmid constructions) were microinjected into Vero cells. The distribution of fusion proteins was determined 18 hr later by indirect immunofluorescence using a mixture of four monoclonal antibodies raised against E. coli β-galactosidase followed by rhodamine-labeled rabbit anti-mouse IgG. No background fluorescence was seen in uninjected cells. Cells expressing fusion proteins either exhibited fluorescence of similar intensity in both the cytoplasm and the nucleus (N/C) or solely nuclear fluorescence (N) (see Figure 6).

bacterial origin, E. coli β-galactosidase appears intrinsically able to localize, at least partially, to the nucleus of higher eucaryotic cells.

A similar distribution between cytoplasm and nucleus was seen for the fusion protein that includes 116 amino-terminal residues of SV40 large T. An additional 19 amino acids (up to 135), however, led to a strikingly different distribution. All of the fusion protein was now seen in the nucleus with no evidence of cytoplasmic staining (Figures 5 and 6a). An equivalent fusion protein that contained Thr in place of Lys-128 was, by contrast, again found to be distributed between nucleus and cytoplasm (Figures 5 and 6b). From this it appears that the conversion of a fusion protein consisting predominantly of β-galactosidase from a partially nuclear to a totally nuclear phenotype can be achieved by sequences derived from SV40 large T, provided Lys-128 and the immediately adjacent residues are present.

The possibility that the same effect could be achieved by the sequence around Lys-128 in isolation was tested by inserting a small region of the SV40 large T gene just downstream of the normal large T initiation codon and before the codon for amino acid 8 of β-galactosidase (Figure 5). Again, the fusion protein was entirely nuclear if this short sequence (amino acids 127 to 147 of large T) included Lys at position 128 but was seen in both the nucleus and the cytoplasm if Thr replaced Lys-128 (Figures 5, 6c, 6d).

The β-galactosidase fusion proteins clearly demonstrate the ability of a very short amino acid sequence including Lys-128 to promote the nuclear accumulation of that fraction of the protein normally found in the cytoplasm. However, β-galactosidase is somewhat unsatisfactory in this assay because the native protein (or, strictly, slight variants thereof), is not completely excluded from the nucleus.

Localization of Pyruvate Kinase Fusion Proteins

A series of experiments similar to those described above were conducted using chicken muscle type M1 pyruvate kinase in place of β-galactosidase. This protein shares some of the favorable characteristics listed above for β-galactosidase: the monomeric protein is reasonably large (theoretical M_r = 58,000; Lonberg and Gilbert, 1983), the active form is tetrameric (Cardenas et al., 1975), and antibodies are available that recognize the protein. Additionally, the crystal structure of cat muscle pyruvate kinase has been determined (Stuart et al., 1979) and reveals that the amino terminus of the protein is exposed and therefore provides a suitable environment for a putative nuclear location signal.

A cDNA encoding chicken muscle pyruvate kinase (Lonberg and Gilbert, 1983) was engineered so as to include the SV40 early region polyadenylation sequence following its normal termination codon and a unique Eco RI site 17 amino acids distal to the normal amino terminus of the protein (section 6 in Experimental Procedures). To this Eco RI site were joined the SV40 early promoter and sequences encoding the first 4, 116, or 135 amino acids of SV40 large T. The subcellular location of the pyruvate kinase fusion proteins was examined by indirect immunofluorescence following microinjection of DNA into Vero cells, using an antibody raised against chicken muscle type M1 pyruvate kinase (Lonberg and Gilbert, 1983). The fusion proteins that included 4 or 116 amino acids derived from large T (Figure 7) were detected throughout the cytoplasm but were absent from the nucleus of injected cells. In stark contrast, the fusion protein that included large T amino acids up to 135 was seen only in the nucleus of injected cells (Figure 6e). Only a cytoplasmic signal was seen if a Thr residue replaced Lys-128 in this construct (Figure 6f).

The clear inference from these observations is that the same short sequence derived from SV40 large T that accentuates the nuclear localization of β-galactosidase fusion proteins is responsible for changing the normal subcellular location of another protein, pyruvate kinase from the cytoplasm of the cell to the nucleus. This prediction was then tested directly and, at the same time, approximate limits were set on the length of large T amino

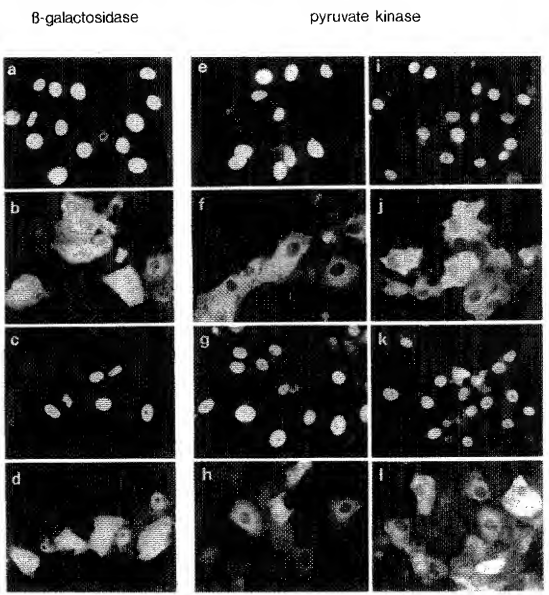

Figure 6. Subcellular Location of β-Galactosidase and Pyruvate Kinase Fusion Proteins

Vero cells were microinjected with plasmid DNAs encoding fusion proteins illustrated in Figure 5 (β-galactosidase fusions) and Figures 7 and 8 (pyruvate kinase fusions). Eighteen hours later the cells were fixed and the fusion proteins were visualized by immunofluorescence microscopy. Micrographs (a) to (d) represent β-galactosidase fusions, visualized with a mixture of monoclonals directed against β-galactosidase followed by rhodamine-conjugated rabbit anti-mouse IgG, as follows. (a) L27-BGAL, (b) d10/L27-BGAL, (c) L7R-BGAL, (d) d10/L7RH-BGAL (see Figure 5). Micrographs (e) to (l) show pyruvate kinase fusions stained with a serum raised in rabbits against purified chicken muscle pyruvate kinase, and counter-stained with fluorescein-labeled goat anti-rabbit serum, as follows. (e) L27-PK, (f) d10/L27-PK, (g) XR4-PK, (h) XRc10/L27-PK, (i) XR30-PK, (j) XR22-PK, (k) m30-PJ-B, (1) xm30-PK (see Figures 7 and 8).

acid sequence that suffices to direct pyruvate kinase to the nucleus.

Large T Amino Acid Sequences Sufficient for Nuclear Localization of Pyruvate Kinase

The method used to construct plasmids encoding pyruvate kinase fusion proteins with short putative nuclear location signals at their amino termini was, in principle, the same as that used for β-galactosidase. End points, defined by restriction enzyme cleavage sites, were created in each of the three possible reading frames both following the amino-terminal five residues of large T and preceding residue 17 of pyruvate kinase. Thus any small fragment of known nucleotide sequence could be inserted between appropri-

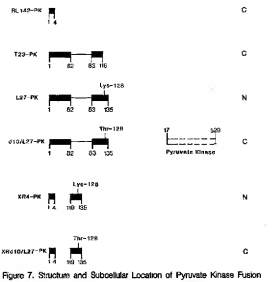

Figure 7. Structure and Subcellular Location of Pyruvate Kinase Fusion Proteins

Plasmids encoding a variable number of amino acid residues derived from the amino terminus of large T fused to pyruvate kinase (see sections 6 and 7 of Experimental Procedures for their construction) were microinjected into Vero cells. The distribution of fusion proteins was determined 18 hr later by indirect immunofluorescence using serum from rabbits immunized with chicken muscle type VI pyruvate kinase followed by fluorescein-labeled goat antirabbit antibody. Although some background fluorescence was seen in uninjected cells, the fusion proteins were clearly either cytoplasmic (C) or nuclear (N) (see Figure 6, e–i).

ate end points such that the normal large T and pyruvate kinase reading frames were maintained across both fusion junctions (details in Section 7 in Experimental Procedures). The small fragment was generated in the first instance by cloning the Hinf I (4442)–Eco RI (at 4413) fragment encoding either Lys or Thr at position 128. Expression of these proteins in Vero cells showed that amino acids 119–135 of SV40 large T sufficed to promote complete nuclear accumulation of the pyruvate kinase fusion protein, whereas a similar sequence including Thr in place of Lys-128 showed no such capability (Figures 6g, 6h, and 7).

The length of large T amino acid sequence preceding residue 135 was then progressively reduced by exonuclease digestion of the wild-type Hinf I (4442)–Eco RI (4413) fragment before insertion into the appropriate pyruvate kinase fusion vector. The amino acid sequences of these fusion proteins and their subcellular location are shown in Figure 8. SV40 large T amino acids from 126 to 135 (XR30 PK, XR27 PK) promoted efficient nuclear accumulation of pyruvate kinase (Figure 6i) while 132 to 135 (XR22 PK) did not discernibly alter the cytoplasmic location characteristic of native pyruvate kinase (Figure 6j). We did not, however, observe a sharp transition between these two extremes. XR12 PK, which encodes large T amino acids 129 to 135 gave rise to a protein that distributed between nucleus and cytoplasm. Exact interpretation of the large T amino acid sequence sufficient to act as a signal is complicated by the contribution that may be made to this activity by adjacent amino acids encoded by linker DNA sequences. Note in particular that the residue preceding 129 in XR12 PK is Arg, which is the only amino acid that has so far been found to be able to replace Lys-128 in large T without totally preventing accumulation in the nucleus (W. H. Colledge, personal communication).

The distal end point of the nuclear location signal was refined further by taking advantage of an adjacent Mbo II site in order to introduce an Eco RI site following the nucleotides encoding Val-132 (Section 7 of Experimental Procedures). Even this short signal extending from amino acid 126 to 132 was found to be sufficient for nuclear accumulation (Figures 6k and 8). Exonuclease digestion from the Mbo II cleavage site allowed the generation of one more fusion protein that included large T sequences from residue 126 to 130 and which was found to localize exclusively in the cytoplasm (Figures 6l and 8). Thus, taking into account possible contributions from linker sequences (see underlined residues in Figure 8), we believe that the minimal sequence of SV40 large T amino acids that suffices to promote efficient nuclear accumulation of pyruvate kinase is most probably ^{126}Pro-Lys-Lys-Lys-Arg-Lys-Val^{132}.

Discussion

Isolation of a Nuclear Location Signal

Evidence has been presented to show that a short sequence of amino acids around Lys-128 is required for the normal nuclear localization of SV40 large T and can suffice when appropriately integrated into the structure of cytoplasmic proteins to promote their accumulation into the nucleus. This nuclear location signal fails to function in its normal position in large T, or at the amino terminus of large T, β-galactosidase, or pyruvate kinase fusion proteins when Lys-128 (of large T) is replaced by a Thr residue. The signal therefore satisfies the condition of specificity that would be expected in order that distinct populations of proteins be maintained in cytoplasmic and nuclear compartments.

It seems likely that the sequence of amino acids including Lys-128 itself constitutes a structure that is directly recognized as a nuclear location signal rather than inducing changes in the conformation of the rest of the polypeptide chain to which it is attached, since it has been shown to direct three different proteins to the nucleus.

Sequence of Nuclear Location Signal

The length of amino acid sequence sufficient to act as a nuclear location signal for pyruvate kinase has been determined in this study to be approximately seven amino acids (126–132). This agrees reasonably well with estimates of the sequence requirements for the nuclear location of large T, as judged by the properties of both deletion and point mutants (Kalderon et al., 1984, and unpublished results).

We already know from an analysis of the subcellular location of point mutants of large T that variations in the

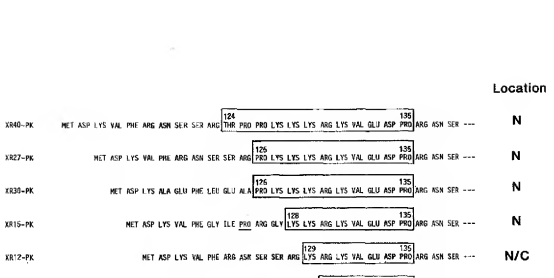

Figure 8. Definition of the Minimal Nuclear Location Signal

The entire amino acid sequence preceding the junction with pyruvate kinase sequences is shown for fusion proteins that include only short regions of large T sequence. The sequence that is directly derived from the region around Lys-128 in large-T is boxed in each case. Residues encoded by the linkers that surround this sequence are underlined where they are, fortuitously, identical with the amino acid residue found in the equivalent position in large T. Most fusion proteins were either clearly cytoplasmic (C) or nuclear (N) but one indicated as N/C was detected in both compartments.

minimal nuclear location signal can be accommodated without complete loss of function and that the residue occupying position 128 is of prime importance (Kalderon et al., 1984 and unpublished results). Definition of the exact parameters to which this sequence must conform should help to elucidate the structural basis of the action of this signal and help to identify the occurrence of potentially active signals in proteins other than SV40 large T. In preliminary searches of other protein sequences we have found several nuclear proteins (e.g., adenovirus 72 kd DNA-binding protein, SV40, and polyoma capsid proteins) that include similar sequences to the large T prototype but only one (BK virus large T) that is identical. The functional activity of such homologous sequences may depend not only on their composition but also on their location within the protein. It will therefore be instructive to examine the efficacy of the large T signal when integrated at different sites within a protein of known tertiary structure.

Is the Nuclear Location Signal a Nuclear Transport Signal?

The existence of nuclear location signals was originally postulated to account for the transport of a specific subset of large proteins to the nucleus (De Robertis et al., 1978). Although the nuclear location signal described here may fulfil this function we have defined its activity only by observing its effect on nuclear accumulation. Further experiments outlined below designed to test whether the signal accentuates the nuclear concentration of proteins by promoting transport are less conclusive.

SV40 large T and the fusion proteins described above are sufficiently large that they would not be expected to accumulate in the nucleus as rapidly as observed, in the absence of a specific transport mechanism. An effect on transport is also indicated by the reduced rate of nuclear accumulation of large T variants with defective signals (Kalderon et al., 1984). Also, although it is not clear whether an association with a fixed nuclear component contributes to the accumulation of wild-type large T in the nucleus, it is certain that the failure of d10 large T to accumulate in the nucleus is not due to an inability to bind to SV40 origin DNA or the cellular protein, p53, since we know that it retains these activities (Paucha et al., submitted).

Thus, although not compelling, the available evidence favors the possibility that the Lys-128 nuclear location signal acts by promoting transport to the nucleus of pro-

teins that are otherwise too large to penetrate the nuclear envelope.

Experimental Procedures

Plasmid Constructions

1) Construction of Deleted, Duplicated, and Truncated Versions of Large T in Wild-Type and d10 (Lys-128→Thr) Backgrounds

Plasmids encoding large T variants were generated by using native restriction sites of the large T gene and derivatives (Kalderon and Smith, 1984) of the wild-type gene in which a 10 bp Eco RI linker has been inserted at the positions shown in Figure 1.

Large T derivatives that encode two adjacent copies of the sequence from 127 to 135 separated by the coding information of an Eco RI linker were constructed by ligation of the large Bam HI (2533)–Eco RI (4413) fragment of L27-S26 or d10/L27-S26 to the small Bam HI (2533)–Eco RI (4441) fragment of T23-L7 or T23-d10/L7.

A synthetic self-complementary oligonucleotide of sequence CTAGT-TAACTAG that includes an Hpa I site and encodes a termination codon in all three reading frames, was inserted at the Pvu II site of the wild-type gene clones in plasmid pPVU-0 (Kalderon et al., 1982) and of the mutant plasmid, d10 (Kalderon et al., 1984). The termination linker was also inserted at the Hpa I site (3733) in both wild-type and d10 plasmids after addition of an Eco RI linker to this site. Finally, the termination linker was inserted into the Eco RI sites of RL148, RL67, and RL50, and for RL82 and RL50 was transferred to a d10 background by exchanging the Pvu II–Bam HI fragment. A termination codon was created at the Hind III site at 4002 as for the mutant dl 1001 (Rundell et al., 1977) by excision of the Hind III 4002–3476 fragment from pPVU-0 and from d10, creating the sequence ^{4002}AAG CTT TAA^{3999}. Plasmids, referred to as EXON2 and d10/EXON2, theoretically encoding large T derivatives which initiate at Met-102 and which certainly include no information from the first exon of large T, were made by addition of an Eco RI linker to the Taq I (4739) sites of pPVU-0 and d10, respectively, and subsequent ligation of this site to the Eco RI site of plasmid RL16, thus deleting nucleotides 2193–4739.

2) Conversion of Linker Insertion/Deletion Mutants from Wild-Type to d10 Background

The linker insertion/deletion mutants, T23-L7, L27-S26, and L26-S24 were each combined with the d10 point mutation by the following method. A heteroduplex was formed between L27-S26 and d10 as described previously (Kalderon et al., 1982) and was treated with the minimal activity of S1 nuclease (about 10 U/μg DNA/20 μl) required to nick one or other but not both of the single-stranded loops in this heteroduplex. The criterion used in practice was the conversion of less than one quarter of the circular heteroduplex to a linear form. Following transfection into E0676? the nicked strands are removed and the gap repaired by copying the intact strand. Following resolution of the strands of the heteroduplex, the resulting molecules consist of one or other of the parental plasmids, wild-type plasmid, or the desired recombinant. The colonies were screened for the required plasmids containing both an Eco RI site in the appropriate position and the d10 sequence at position 4435, by hybridization of Southern blots (Southern, 1975) of Pst I–Eco RI-digested plasmid DNA to oligonucleotide probe of wild-type sequence from 4441–4423 at increasing temperature (Kalderon et al., 1984). Approximately one third of plasmids were found to be of the required structure. This was confirmed by DNA sequence determination for plasmids T23-d10/L7 and d10/L27-S26. This method is simple, of general application to the combination of deletion and point mutations irrespective of the distribution of restriction sites and has also been used to combine two nonoverlapping deletions.

Plasmid T50-d10/L7 was made by replacing the ~1800 bp Bam HI (2533)–Eco RI (4441) fragment of T50-L7 with the equivalent fragment from T23-d10/L7.

3) Vectors Encoding Amino-Terminal Regions of Large T

The Bam HI (2533)–Eco RI (5150) fragment of RL142 (Figure 1) includes the origin of plasmid DNA replication, the β-lactamase gene, the SV40 early promoter, and encodes the first four amino acid residues that are common to SV40 small t and large T. The exact position of the Eco RI site in RL142 was refined by mild S1 nuclease digestion after cleavage with Eco RI, followed by the addition of another (10-mer) Eco RI linker to the newly created end points. Plasmids SRL40 and SRL90 produced in this way contain an Eco RI linker abutting nucleotides 5154 and 5149, respectively (where nucleotide 5149 is now a C residue deriving from the Eco RI linker in RL142). Thus DNA fragments ligated to the Eco RI site can be read in each of the three possible reading frames by using the appropriate vector. RL142, SRL40, or SRL90. A similar set of end points was derived from RL142, SRL30, and SRL40 by the addition of an Xho I linker (CCTCGAGG) to the filled-in, unique Eco RI site, forming the plasmids XRL142, XSRL30, and XSRL40. Vectors that encode larger amino-terminal segments of large T were made from the Bam HI (2533)–Eco RI fragments of the mutants T23-L7, L27-S26, and d10/L27-S26, each of which include a 10 bp Eco RI linker.

4) Relocation of a Short Sequence around Lys-128 to the Amino Terminus of Large T

Small fragments of the large T gene encoding either amino acid residues 119–135 or 127–147 with, in each case, either Lys or Thr at position 128 were constructed from the plasmids L27-d10/L7 and d10/L27-L7 (Figure 3a). These plasmids were each digested to completion with Hinf I, a Hind III linker (CCAAGCTTGG) was added to all free ends, and, after digestion with Hind III and Eco RI, the mixture of fragments was ligated to the large Hind III–Eco RI fragment of pBR328. The two required products of each ligation were distinguished by hybridization at different temperatures to an oligonucleotide of wild-type sequence from 4441–4423 (corresponding to amino acid residues 127–132) and were subsequently sequenced from the Eco RI site. Purified Eco RI (4441)–Hind III (4376) fragments derived from two of these plasmids (L7RH and d10/L7RH) were ligated to the Bam HI (2533)–Eco RI (5149) fragment of SRL30 and the Hind III (5317)–Bam HI (2533) fragment of S11-S33 to form chimeras in which the region around Lys-128 has been transposed to the amino terminus of large T. The amino acid sequences of the chimeras is shown in Figure 3b.

5) Construction of β-Galactosidase Fusion Vectors

The plasmids pSK107 and pMC1871 (Casadaban et al., 1983) each include the majority of the E. coli *lac Z* gene encoding amino acid residue 6 and onwards of β-galactosidase. A Bam HI site is present just after the β-galactosidase termination codon in pMC1871 and the polylinker sequences that precede the *lac Z* gene in pSK107 are suitably disposed for fusion with SV40 large T coding sequences at either the Eco RI or Hind II sites. The plasmid BGAL1, which contains polylinker and the SV40 early region polyadenylation signal flanking the *lac Z* gene, was constructed in a three-fragment ligation using the smaller Pst I–Cla I fragment from pSKS107, the smaller Cla I–Bam HI fragment of pMC1871, and the somitor Bgl II–Pst I fragment of RL18-BGL2666. The latter plasmid was constructed by conversion of the Hpa I site (2666) of RL18 (Kalderon and Smith, 1984) to a Bgl II site using synthetic linkers.

Suitable fragments, purified by agarose gel electrophoresis, were generated from this vector by partial Bam HI–Eco RI digestion for fusion with Bam HI–Eco RI fragments of vectors encoding the amino-terminal regions of large T (section 3, Experimental Procedures) to form SRL30-BGAL, T23-BGAL, L27-BGAL, and d10/L27-BGAL (Figure 5). The Hind III–Bam HI fragment of BGAL-1 was used to construct the plasmids L7RHBGAL and d10/L7RHBGAL (Figure 5) by simultaneous ligation to the Bam HI (2533)–Eco RI (5149) fragment of SRL30 (section 3, Experimental Procedures) and the Eco RI (4441)–Hind III (4376) fragment from either L7RH or d10/L7RH (section 4, Experimental Procedures).

6) Construction of Pyruvate Kinase Fusion Vectors

The SV40 early region polyadenylation signal, derived from the plasmid RL18;BGL 2666, was added to the 3′ end of the cDNA of chicken muscle type M1 pyruvate kinase contained in plasmid pPK300 (constructed by N. Lonberg by joining plasmids pPK201 and pPK101 at their common BstE II site; Lonberg and Gilbert, 1983). This was achieved by ligation of Bgl II–Eco RI fragments following the conversion of the unique Pvu II and Apa I sites of pPK300 to Eco RI and Bgl II sites, respectively. Other vectors, RL18-PK10 and RL18-PK12, were constructed in an analogous fashion but using Eco RI linkers of 10 bp (CGGAATTCCG) and 12 bp (CCGGAATTCCGG), respectively, rather than 8 bp (GGAATTCC) to convert the Pvu II site of pPK300 to an Eco RI site. An Eco RI–Bam HI fragment that contains codon 17 onwards of pyruvate kinase followed by a termination codon and a polyadenylation signal was purified from these plasmids and ligated to Bam HI–Eco RI fragments of the plasmids RL142, T23-L7, L27-S26, or d10/L27-S26, which encode a variable number of amino-

terminal residues of large T prior to the Eco RI site, to form the pyruvate kinase gene fusions shown in Figure 7.

7) Construction of Pyruvate Kinase Chimeras That Include Minimal Sequences from SV40 Large T

The small Hind III–Eco RI fragment of plasmid L27HR (section 4 of Experimental Procedures) that includes SV40 DNA sequences 4459–4413 was shortened by Bal 31 exonuclease digestion (1 U/4 μg DNA/20 al for 10 min at 30°C) from the Hind III site, followed by the addition of an Xho I linker (CCTCGAGG) to the digested ends and subsequent circularization. The positions of the Xho I linker were determined for several products (XR4, 9, 12, 15, 22, 24, 27, 30, 40) by sequencing the small Bam HI–Eco RI fragment of these plasmids from the Eco RI site. Purified Xho I–Eco RI fragments from these plasmids were incorporated into a suitable pyruvate kinase fusion protein vector by simultaneous ligation to the Bam HI–Xho I fragment of XRL142, XSRL30, or XSRL40 (section 3, Experimental Procedures) and the Bam HI–Eco RI fragment of RL18.PK8, RL18.PK10, or RL18.PK12 (section 6 of Experimental Procedures). A similar short sequence derived from d10/L27HR was also incorporated into a pyruvate kinase vector in similar fashion after conversion of the Hind III site of d10/L27HR to an Xho I site.

The Xho I–Eco RI of XR30 was reduced further in size, keeping the Xho I site fixed in the following way. XR30 was digested with Mbo I and a portion was treated with Bal 31 exonuclease. In each case, the shortened 450 bp fragment containing SV40 sequences adjacent to the Xho I site was purified by agarose gel electrophoresis after the addition of either 8 bp, 10 bp, or 12 bp Eco RI linkers. Subsequent to digestion with Eco RI and Xho I, the resultant mixture of fragments was used directly to clone the required shortened SV40 fragment into a pyruvate kinase fusion vector by three-fragment ligation, as described above. The DNA sequence from the Hind III (5171) site contributed by SV40 large T coding sequence up to and beyond the start of pyruvate kinase coding sequences (which includes all ligation junctions) was verified for all constructs shown in Figures 7 and 8.

Subcellular Localization of Proteins

The procedures used to isolate transformed cell lines and to induce transient expression of proteins by needle microinjection of DNA as well as the indirect immunofluorescent staining procedure used have been described previously (Kalderon et al., 1984). The specificities of all antibodies used have been demonstrated previously: PAb419 and PAb423 (Harlow et al., 1981); mouse anti-β-galactosidase antibodies (a kind gift of Dr. J. A. Partiadis, University of Massachusetts, Amherst); rabbit anti-chicken muscle type M1 pyruvate kinase (Lonberg and Gilbert, 1983); FITC-conjugated rabbit anti-mouse IgG (a kind gift of Dr. Roger Morris, National Institute for Medical Research, London); FITC-conjugated rabbit anti-hamster and goat anti-rabbit serum (supplied by Nordic Ltd.) and were verified in the experiments described here by the level of background staining seen in uninjected cells.

Acknowledgments

We thank Dr. M. J. Casadaban (University of Chicago) for plasmids containing the *lac Z* gene, Dr. J. A. Partiadis (University of Massachusetts, Amherst) for monoclonal antibodies to β-galactosidase, Dr. N. Lonberg (Biogen Inc., Cambridge) for the pyruvate kinase cDNA clone, and Dr. N. Lonberg and Dr. D. Hart (Hershey Medical Center, University of Pennsylvania) for antiserum raised against chicken muscle type M1 pyruvate kinase. We also thank Drs. B. A. Costa and A. F. Markham for the design and synthesis of the termination linker oligonucleotide, Ms. Yvette Hartley for technical assistance, Mrs. Lydia Pearson for typing the manuscript, and Joe Brock, Mike Tatham, and the staff of the Photography Department for help with illustrations.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 26, 1984; revised October 17, 1984

References

Bickel, G. (1982). Intracellular protein topogenesis. Proc. Nat. Acad. Sci. USA *77*, 1496–1500

Bonner, W. M. (1975). Protein migration into nuclei. II. Frog oocyte nuclei accumulate a class of microinjected oocyte nuclear proteins and exclude a class of microinjected oocyte cytoplasmic proteins. J. Cell Biol. *64*, 431–437.

Cardenas, J. M., Blazhy, E. G., Ceccotti, P. L., and Dyson, R. D. (1975). Properties of chicken skeletal muscle pyruvate kinase and a proposal for its evolutionary relationship to the other avian and mammalian isozymes. Biochemistry *14*, 2247–2252.

Casadaban, M. J., Martinez-Arias, A., Shapira, S. K., and Chou, J. (1983). β-Galactosidase gene fusions for analyzing gene expression in Escherichia coli and yeast. Meth. Enzymol. *100*, 293–308.

Craig, N., and Perry, R. P. (1971). Persistent cytoplasmic synthesis of ribosomal proteins during the selective inhibition of ribosomal RNA synthesis. Nature *229*, 75–80.

De Robertis, E. M. (1983). Nucleocytoplasmic segregation of proteins and RNAs. Cell *32*, 1021–1025.

De Roberts, E. M., Longthorne, R. F., and Gurdon, J. B. (1978) Intracellular migration of nuclear proteins in *Xenopus* oocytes. Nature *272*, 254–256.

Dingwall, C., Sharnick, S. V., and Laskey, R. A. (1982). A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. Cell *30*, 449–458

Feldherr, C. M. (1975). The uptake of endogenous proteins by oocyte nuclei. Exp. Cell Res. *93*, 411–419.

Fowler, A. V., and Zabin, I. (1977). The amino acid sequence of β-galactosidase of *Escherichia coli*. Proc. Nat. Acad. Sci. USA *74*, 1507–1510

Gurdon, J. B. (1970). Nuclear transplantation and the control of gene activity in anima development. Proc. Royal Soc. Lond. *B 176*, 303–314.

Hall, C. V., Jacob, P. E., Hingold, G. M., and Lee, F. (1983). Expression and regulation of *Escherichia coli lac Z* gene fusions in mammalian cells. J. Mol. App. Genet. *2*, 101–109.

Hall, M. N., Hereford, L., and Hershowitz, I. (1984). Targeting of Escherichia coli β galactosidase to the nucleus in yeast. Cell *36*, 1057–1065.

Harlow, E., Crawford, L. V., Pim, D. C., and Williamson, N. M. (1981). Monoclonal antibodies specific for simian virus 40 tumor antigens. J. Virol. *39*, 861–869.

Harris, J. R. (1981). In Electron Microscopy of Proteins, Vol. 1. (New York: Academic Press).

Kalderon, D., and Smith, A. E. (1984). *In vitro* mutagenesis of a putative DNA binding domain of SV40 large-T. Virology, in press.

Kalderon, D., Costa, B. A., Ely, B. K., and Smith, A. E. (1982). Deletion-loop mutagenesis: a novel method for the construction of point mutations using deletion mutants. Nucl. Acids Res. *10*, 5161–5171.

Kalderon, D., Richardson, W. D., Markham, A. F., and Smith, A. E. (1984). Sequence requirements for nuclear localisation of SV40 large-T antigen. Nature *311*, 33–38.

Lanford, R. E., and Butel, J. S. (1980). Biochemical characterization of nuclear and cytoplasmic forms of SV40 tumor antigens encoded by parental and transport-defective mutant SV40-adenovirus 7 hybrid viruses. Virology *105*, 314–327.

Lanford, R. E., and Butel, J. S. (1984). Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. Cell *37*, 801–813.

Lonberg, N., and Gilbert, W. (1983). Primary structure of chicken muscle pyruvate kinase mRNA. Proc. Nat. Acad. Sci. USA *80*, 3661–3665.

Paine, P. L., and Feldherr, C. M. (1972). Nucleocytoplasmic exchange of macromolecules. Exp. Cell Res. *74*, 81–98.

Paine, P. L., Moore, L. C., and Horowitz, S. B. (1975). Nuclear envelope permeability. Nature *254*, 109–114.

Peters, R. (1984). Nucleocytoplasmic flux and intracellular mobility in single hepatocytes measured by fluorescence microphotolysis. EMBO J. *3*, 1831–1836.

Rundell, K., Collins, J. K., Tegtmeyer, P., Ozer, H. L., Lai, C. J., and Nathans, D. (1977) Identification of simian virus 40 protein A. J. Virol. *21*, 636–646.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517.

Stuart, D. I., Levine, M., Muirhead, H., and Stammers, D. K. (1979). Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 Å. J. Mol. Biol. 134, 109–142.

Tjian, R., Fey, G., and Graessmann, A. (1978). Biological activity of purified simian virus 40 T-antigen proteins. Proc. Nat. Acad. Sci. USA 75, 1279–1283.

Tooze, J. (1981). Molecular biology of tumor viruses. Part 2. DNA tumor viruses, revised edn. 2. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Unwin, P. N. T., and Milligan, R. A. (1982). A large particle associated with the perimeter of the nuclear pore complex. J. Cell Biol. 93, 63–75.

Wallenfels, K., and Weil, R. (1972). In The Enzymes, VII, 3rd edition. P. D. Boyer, ed. (New York: Academic Press), pp. 617–663.

Watson, M. E. E. (1984). Compilation of published signal sequences. Nucl. Acids Res. 12, 5145–5164.

Wu, R. S., and Warner, J. R. (1971). Cytoplasmic synthesis of nuclear proteins: kinetics of accumulation of radioactive proteins in various cell fractions after brief pulses. J. Cell Biol. 51, 643–652.

Yamazumi, M., Uchida, T., Okada, Y., Furusawa, M., and Mitsui, H. (1978). Rapid transfer of non-histone chromosomal proteins to the nucleus of living cells. Nature 273, 782–784.

HOS cells and as the chemistry of the interaction of simple methylating carcinogens such as MNNG is particularly well defined[46], a comparison of the *met* gene from MNNG–HOS cells with the parental HOS cell gene may provide a model system with which to study the mechanisms of transformation of human cells by this carcinogen.

In at least one *N*-nitroso-*N*-methylurea (NMU) induced rat mammary tumour, activation of the *ras*^{H} transforming gene involves a simple G→A point mutation[3]. Notably, this is precisely the mutation that NMU would be predicted to induce if it reacted directly with the transforming gene. It will be interesting to see whether the precise method of activation of this transforming gene and of other transforming genes detected in chemically transformed cells also reflects the particular chemicals used to induce transformation. Further characterization of the *met* transforming gene and its comparison with the non-transforming allele may provide an interesting insight into one of the ways in which chemical carcinogens can activate proto-oncogenes in human cells.

We thank Dr M. Wigler for the gift of the *ras*^{N} clone, Drs C. Der, A. Diamond, G. M. Cooper and M.-A. Lane for providing the *Blym*, *sis*, *myb*, *src*, *erb* and *rel* clones, Dr B. Vennstrom for the human *erbB* clone, Dr Barbara Knowles for the 53-87-3 c136 is c12 sub7 cell lines, Dr David Patterson for 706B6-40 c117 DNA, and Debbie Dugan and Josephine Romano for technical assistance. This work was in part supported by the NCI, DHHS, under contract N01-CO-23909 to G.V.W. and NIH grants CA16685, CA21124 and HD17561 to K.H. and C.M.C.

Received 7 March; accepted 6 June 1984.

1. Balmain, A. & Pragnell, I. B. *Nature* **303**, 72–74 (1983).
2. Balmain, A., Ramsden, M., Bowden, G. T. & Smith, V. *Nature* **307**, 658–660 (1984).
3. Sukumar, S., Notario, V., Martin-Zanca, D. & Barbacid, M. *Nature* **306**, 658–661 (1983).
4. Sukumar, S. *et al. Science* **223**, 1197–1199 (1984).
5. Eva, A. & Aaronson, S. A. *Science* **220**, 955–956 (1983).
6. Parada, L. F. & Weinberg, R. A. *Molec. cell. Biol.* **3**, 2298–2301 (1983).
7. Guzmano, I., Caleada, P., Mayer, A. & Pellicer, A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 202–205 (1984).
8. Henry, S. A. *Br. med. Bull.* **4**, 389–401 (1947).
9. USPHS Publ. (HEW) 79-50066 (US Dept of Health, Education and Welfare, Washington, DC, 1979).
10. Renwick, A. G. & Drasm, B. S. *Nature* **263**, 234–235 (1976).
11. Kakunaga, T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1334–1338 (1978).
12. Rhim, J. S. *et al. J. natn. Cancer Inst.* **55**, 1291–1294 (1975).
13. Rhim, J. S., Park, D. K., Arnstein, P., Huebner, R. J. & Weinburger, E. K. *Nature* **256**, 751–753 (1975).
14. Diffie, J. A. *J. natn. Cancer Inst.* **70**, 3–8 (1983).
15. Bisle, D. G., Cooper, C. S., Oskarsson, M. K., Eader, L. A. & Vande Woude, G. F. *Science* **218**, 1122–1125 (1982).
16. Cooper, C. S. *et al. Cancer Res.* **44**, 1–10 (1984).
17. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
18. Jelinek, W. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1398–1402 (1980).
19. Shimizu, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2112–2116 (1983).
20. Hall, A., Marshall, C. J., Spurr, N. K. & Weiss, R. *Nature* **303**, 376–400 (1983).
21. Diamond, A., Cooper, G. M., Ritz, J. & Lane, M. A. *Nature* **305**, 112–116 (1983).
22. Chang, E. H., Gonda, M. A., Ellis, R. W., Scolnick, E. M. & Lowy, D. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4648–4652 (1982).
23. Watson, R., Oskarsson, M. & Vande Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4078–4082 (1982).
24. Dalla Favera, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 6497–6501 (1982).
25. Heisterkamp, N., Groffen, J. & Stephenson, J. R. *J. molec. appl. Genet.* **2**, 57–68 (1983).
26. Josephs, S. F., Dalla Favera, R., Gelmann, E. P., Gallo, R. C. & Wong-Staal, F. *Science* **219**, 503–505 (1983).
27. Groffen, J., Heisterkamp, N., Grosveld, F., Van de Ven, W. & Stephenson, J.-R. *Science* **216**, 1136–1138 (1983).
28. Dalla-Favera, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 7824–7827 (1982).
29. Croce, C. M. & Koprowski, H. *J. exp. Med.* **159**, 1350–1393 (1974).
30. Knowles, B. A. *et al. J. exp. Med.* **145**, 514–526 (1977).
31. Croce, C. M., Aden, D. & Koprowski, H. *Science* **190**, 1200–1202 (1975).
32. Cohen, R. V., Rice, N. R. & McAllister, R. M. *Gene Analysis Tech.* **1**, 23–33 (1984).
33. Spurr, N. K. *et al. EMBO J.* **3**, 159–163 (1984).
34. Vennstrom, B., Fanshier, L., Moscovici, C. & Bishop, J. M. *J. Virol.* **36**, 575–585 (1980).
35. Jansson, M., Philipson, L. & Vennstrom, B. *EMBO J.* **2**, 361–365 (1983).
36. Rowley, J. D. & Testa, J. R. *Adv. Cancer Res.* **36**, 103–148 (1982).
37. Appendix I: *Cancer Genet. Cytogenet.* **11**, 305–363 (1984).
38. Yunis, J. J. *Cancer Genet. Cytogenet.* **11**, 125–137 (1984).
39. Ray, J. H. & German, J. in *Genes, Chromosomes and Neoplasia* (eds Arrighi, F. E., Rao, P. N. & Stubblefield, E.) 351–376 (Raven, New York, 1981).
40. Pulciani, S. *et al. Nature* **300**, 539–542 (1982).
41. Der, C. J. & Cooper, G. M., *Cell* **32**, 201–208 (1983).
42. Perucho, M. *et al. Cell* **27**, 467–476 (1981).
43. Lane, M.-A., Sainten, A. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5185–5189 (1981).
44. Lane, M. A., Sainten, A. & Cooper, G. M. *Cell* **28**, 873–880 (1982).
45. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. *Nature* **300**, 149–152 (1982).
46. Tapoosowsky, E. *et al. Nature* **300**, 762–764 (1982).
47. Tabin, C. J. *et al. Nature* **300**, 143–148 (1983).
48. Marshall, C. P. & O'Connor, P. J. in *Chemical Carcinogens and DNA* Vol. 1 (ed. Grover, P. L.) 111–159 (CRC Press, Boca Raton, Florida, 1979).
49. Hoku, B. & Murray, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3259–3263 (1977).
50. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
51. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
52. Pritchard, A.-M., Lehrach, H., Poustka, A. & Murray, N. *J. molec. Biol.* **170**, 827–843 (1983).

Sequence requirements for nuclear location of simian virus 40 large-T antigen

Daniel Kalderon, William D. Richardson, Alexander F. Markham* & Alan E. Smith

Biochemistry Division, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK
*ICI Pharmaceuticals Division, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

A point mutation in the simian virus 40 large-T gene, which was generated by mixed oligonucleotide mutagenesis and resulted in the conversion of Lys 128 to Thr, produced a large-T antigen that was detected in the cytoplasm but not the nucleus of cells. Deletions within the surrounding sequence Lys-^{128}*Lys-Lys-Arg-Lys-Val-Glu also produce cytoplasmic large-T and define a region of the protein involved in nuclear location.*

THE large-T antigen of simian virus 40 (SV40), molecular weight M_r 94,000, is sufficient to transform established and non-established rodent cells *in vitro* and to induce tumours in newborn hamsters[1]. It is also required for the productive infection of simian cells by SV40[2].

Large-T is a predominantly nuclear protein[3] but is also found in a modified form in the plasma membrane[4,5]. During a productive infection the nuclear form of large-T regulates the level of viral gene expression and stimulates viral DNA replication through an interaction with the viral origin of DNA replication[6–8]. To investigate whether DNA binding by large-T is also involved in transformation we have introduced mutations[9] into a region coding for a putative DNA binding domain of the protein[6,10–13]. Within this domain is a striking tract of five basic amino acids (^{127}Lys-Lys-Lys-Arg-Lys^{131}) just downstream from a cluster of serine and threonine phosphorylation sites[14]. As the sense of the codons for lysine is unaffected by the action of the commonly used mutagen, sodium bisulphite, to mutate this region we devised a modification of oligonucleotide-directed mutagenesis[15,16] that allows the introduction of several defined point mutations in a single experiment.

The large-T protein encoded by one of the mutants generated in this manner contains a threonine residue in place of Lys 128. This mutant large-T, although able to transform Rat-1 cells with wild-type efficiency as measured by focus formation in high serum, accumulates in the cytoplasm rather than the nucleus of

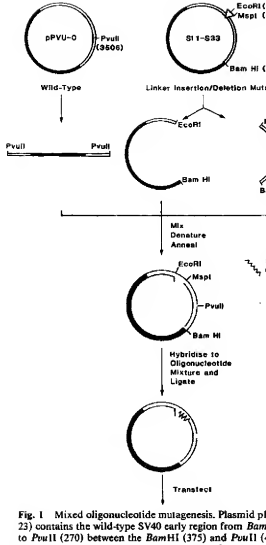

Fig. 1 Mixed oligonucleotide mutagenesis. Plasmid pPVU-0 (ref. 23) contains the wild-type SV40 early region from *Bam*HI (2,533) to *Pvu*II (270) between the *Bam*HI (375) and *Pvu*II (4,676) sites of pBR328 (ref. 37) but with a deletion of 3 base-pairs at the junction of *Pvu*II sites, thus destroying the enzyme recognition site at this point. In the figure, plasmid DNA sequences are shaded whereas the SV40-derived sequences are not. S11-S33 (ref. 9) is a derivative of pPVU-0 that contains a 10 bp *Eco*RI linker (CGGAATTCCG) spanning a deletion between SV40 nucleotides 4441 and 4421. An *Msp*I site is created at the junction of linker and SV40 nucleotide sequences around position 4421. *Msp*I–*Bam*HI and *Bam*HI–*Eco*RI restriction fragments from S11-S33 were purified by agarose gel electrophoresis followed by phenol extraction[9] and were mixed in equimolar proportions with pPVU-0 DNA linearized at its unique *Pvu*II site. After denaturation with alkali and subsequent renaturation[23] about 10–20% of the products migrated as form II (open duplex circles) during agarose gel electrophoresis and were presumed to be circular heteroduplexes of the form shown above. The hybridization products (2.5 μg DNA) were dissolved, after ethanol precipitation in 10 μl of ligation buffer (50 mM Tris-HCl pH 8.0, 7.5 mM $MgCl_2$, 1 mM ATP, 10 mM dithiothreitol (DTT) and 0.02% gelatin) and to this was added 2.5 μg of the oligonucleotide mixture (synthesized as in ref. 38) and phosphorylated by T4 polynucleotide kinase, also in 10 μl of ligation buffer. The components were mixed initially at 60 °C, the temperature was then gradually reduced to 0 °C over a 6 h period, after which additional dithiothreitol (to 5 mM) and T4 DNA ligase (6 units) were added. After 1 h at 0 °C, ligation was continued for 4 h at 15 °C. Some of the ligated material was then used directly for transfection. Alternatively, after phenol extraction and ethanol precipitation aliquots were treated with S_1 nuclease (50 units per 10 μl/0.4 μg DNA for 20 min at 37 °C) before neutralization and transfection. S_1 nuclease treatment reduced the number of transfectants about 10-fold. Parallel experiments showed that significant cleavage of nicked circular plasmid DNA did not occur at the S_1 nuclease concentration used.

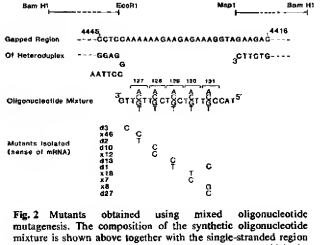

Fig. 2 Mutants obtained using mixed oligonucleotide mutagenesis. The composition of the synthetic oligonucleotide mixture is shown above together with the single-stranded region of the gapped heteroduplex, formed as in Fig. 1, to which the oligonucleotides were hybridized. Note that after hybridization the synthetic oligonucleotides are exactly apposed at their 5′ ends to the double-stranded region of the heteroduplex and can therefore be ligated directly into the heteroduplex. The nucleotide sequence alterations in mutant plamids recovered from the procedure described in Fig. 1 are shown. The inferred alterations to the amino acid sequence of large-T are presented in Table 1. The mutation found in plasmid d3 should not theoretically have been introduced by the oligonucleotide mixture used. It was not studied further because it does not alter the amino acid sequence of large-T.

the transformed cells. The phenotypes of other mutants with lesions in the same area help to define a region around Lys 128 which is required for the nuclear localization of large-T.

Mixed oligonucleotide mutagenesis

Oligonucleotide-directed mutagenesis is currently the method of choice for the introduction of single, precisely defined mutations. However, this method is not generally used for the creation of large families of point mutants because of the labour or expense involved in obtaining a large number of individual synthetic oligonucleotides. In the procedure described here a mixture of oligonucleotides synthesized on a single resin was used as the mutagen.

The mixture of oligonucleotides, each of 20 residues, was synthesized by including an equimolar mixture of all four nucleotides at five selected positions along its length (Fig. 2). The degenerate positions correspond to the second nucleotides in the codons for the basic sequence ^{127}Lys-Lys-Lys-Arg-Lys^{131}, thereby potentially allowing the conversion of each of these residues to any of three different amino acid residues. Thus the oligonucleotide mixture contained 4^5 components, one of which was an exact copy of the wild-type SV40 large-T coding sequence in this region and 15 of which contained a single base change. The oligonucleotide mixture was incorporated into plasmid DNA bearing a cloned wild-type large-T gene by direct ligation into a gapped heteroduplex, formed between wild-type plasmid DNA and two restriction fragments of a linker insertion/deletion mutant derivative of the plasmid (Fig. 1).

After transfection, colony hybridization[17] using a 19-nucleotide long probe of sequence corresponding to the wild-type gene in this region (4,423–4,441: ACCTTTCTCTTCTTTTTGG) was used to detect plasmids bearing mutations within this sequence. Washing filters at increasing temperatures[18] following hybridization at room temperature clearly distinguished between wild-type plasmids in which hybrids were stable up to about 55 °C (in 6×SSC) and others which showed no detectable hybridization beyond about 45 °C. Although this latter class could itself be resolved into two distinct categories because of differences in the stability of hybridization to the oligonucleotide probe, plasmid DNA was isolated from all poorly hybridizing colonies and examined by restriction enzyme mapping. All plasmids for which hybrids were not stable above ~37 °C, contained large

Table 1 Biological properties of mutant large-T proteins produced by mixed oligonucleotide mutagenesis

Wild-type plasmid	127	128	129	130	131	Transformation: Focus formation on Rat-1 cells, 10 μg DNA, Expt	% wt	1 μg DNA, Expt	% wt	Replication: Plasmid DNA	Virus	Location of large-T in transformed Rat-1 cells
X46	Thr					(1)	59	(A)	67	–	–	Nuclear
						(2)	73	(A)	80			
d2	Ile					(3)	28	(A)	135	–	–	Nuclear
						(4)	112	(C)	164	–	–	
d10		Thr				(6)	107	(A)	140	–	–	Cytoplasmic
						(7)	101					
X12		Thr				(2)	16	(A)	83	–	–	Cytoplasmic
						(4)	48					
d13			Thr			(3)	14	(A)	40	+	D*	Mixed
						(5)	23	(B)	61			
d1			Met		Thr	(3)	190	(A)	105	–	–	Mixed
						(7)	96	(E)	378			
X18				Ile		(1)	54	(A)	54	–	–	Mixed
X7				Thr		(2)	33	(A)	31	D†	D‡	Mixed
X8					Arg	(1)	56	(A)	37	+	+	Nuclear
d27					Thr	(3)	24	(D)	74	+	+	Mixed
						(4)	40	(C)	103			

Transformation assays: Plasmid DNA (10 μg or 1 μg+9 μg salmon sperm DNA) was transfected into semi-confluent Rat-1 cells by the calcium phosphate method[20]. Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Dense foci were visible after about 14 days post-transfection and were counted 3–9 days after their first appearance. The numbers above represent the mean of duplicate assays expressed as a percentage of the number of foci induced by the wild-type plasmid, pPVU-0, in the same experiment. The numbers of foci induced by pPVU-0 in these experiments were as follows: Experiment (1) 1,220; (2) 1,150; (3) 446; (4) 500; (5) 464; (6) 694; (7) 634. (A) 170; (B) 94; (C) 50; (D) 76; (E) 62. The deletion mutants RL6 and SS-S32 (ref. 9) which encode a wild-type small-t but severely truncated large-T proteins of 175 and 142 amino acids, respectively, failed to induce any foci in these assays.

Replication assays: Plasmid DNA and viral replication assays were performed as described elsewhere[2]. (–) Indicates that no DNA was detectable in an experiment that would detect 0.1% of wild-type DNA accumulation or that no plaques were detectable within 4 weeks (wild-type plaques are normally visible by 10 days); (+) indicates behaviour indistinguishable from wild-type and (D) (defective) behaviour intermediate between these extremes.

Subcellular localization: Cells derived from transformed foci of Rat-1 cells were grown on glass coverslips, fixed in 4% formaldehyde in phosphate-buffered saline (20 min at room temperature), permeabilized with 1% Nonidet P40 (5 min at room temperature) and incubated with hamster anti-tumour serum followed by fluorescein isothiocyanate-labelled rabbit anti-hamster antibody (Nordic). d10 and X12 transformed cells were also examined using monoclonal antibody PAb423 (ref. 21), followed by rhodamine isothiocyanate-conjugated rabbit anti-mouse antibody. Controls using normal Rat-1 cells and normal hamster serum were included. Normal Rat-1 cells were also generally still present in the population of cells derived from dense foci and thus acted as internal controls for non-specific staining. Immunofluorescence was scored as nuclear, cytoplasmic or mixed. Mixed indicates that three different patterns of staining could be seen in individual cells within a population of transformed cells. Some cells were stained only in the nucleus, others only in the cytoplasm and the rest in both compartments. The most extreme example of this phenotype is d1. Similar localization patterns were seen by staining with the large-T specific monoclonal antibody PAb423 following manual microinjection of each of the mutant plasmids into Vero cells.

* Plaques visible 3 days later than wild-type. † Less than 10% of wild-type DNA accumulation. ‡ Plaques visible 6 days later than wild-type.

deletions (about 2 kb) that included the target for mutagenesis. In contrast, all plasmids for which hybrids were stable to between 40 °C and 45 °C were of wild-type size, indicating the presence of point mutations within the target area. This was confirmed by determining the DNA sequence in this region of all plasmids which formed hybrids of intermediate stability, using the oligonucleotide primed dideoxy procedure applied to double-stranded DNA[19].

Nine of the eleven plasmids sequenced in this way were found to have single point mutations of the wild-type sequence at positions corresponding to the degeneracies in the oligonucleotide mixture (Fig. 2). One mutant plasmid (d1) contained two such alterations and another contained an A to C transversion at a position that should have been invariant (and wild-type) in the oligonucleotide mixture. Two independent isolates of one mutant plasmid (X12; d10) were recovered. In a subsequent experiment using the same oligonucleotide mixture, three further single point mutants were isolated (W. H. Colledge, personal communication). Thus, 12 of the 15 possible point mutants have been obtained. Mutants were recovered corresponding to each of the degenerate positions in the oligonucleotide mixture and examples of each kind of mutagenic change (A → C, G, T) were found. The overall efficiency of production of point mutants was enhanced from 0.7% to 3% by the inclusion of an S_1 nuclease digestion step immediately before transfection, designed to linearize those gapped heteroduplexes to which no oligonucleotide had hybridized. The ease with which large numbers of plasmids can be reliably screened for the presence of point mutations by hybridization to an oligonucleotide of wild-type sequence compensates for the relatively low rate at which mutant plasmids are produced. The genotype of the mutants recovered demonstrates that the mutagenic procedure selects for the incorporation of synthetic oligonucleotides containing single deviations from the wild-type sequence, from a large excess of oligonucleotides with multiple alterations. The principle of mixed oligonucleotide mutagenesis should be applicable to all methods of conventional oligonucleotide-directed mutagenesis currently in use[18].

Transforming activity

The transforming potential of mutant plasmids was assayed by measuring the induction of dense foci on a monolayer of Rat-1 cells following calcium phosphate transfection[20]. The formation of dense foci was large-T dependent, as shown by the failure of deletion mutants that encode a wild-type small-t but severely truncated large-T proteins, to produce dense foci (Table 1). However, all point mutants examined showed a transforming activity similar to wild-type. Although slight differences in specific transforming activity were observed, the latent period before foci became visible was never significantly different from wild-type (13–14 days).

Subcellular localization

Individual foci of transformed Rat-1 cells were picked, the cell population expanded and the subcellular location of large-T assessed by indirect immunofluorescence of permeabilized fixed cells using serum from hamsters bearing SV40-induced tumours as the first antibody.

Rat-1 cells transformed by wild-type large-T, encoded by the plasmid pPVU-0, exhibited a typical nuclear staining with no detectable nucleolar or cytoplasmic staining (Fig. 3*a*). A strikingly different staining pattern was seen for cells transformed by the point mutants, d10 and X12, both of which encode a threonine residue in place of Lys 128 (Fig. 3*c*). The nuclei appear dark against a bright cytoplasmic background. An identical pattern was seen using a monoclonal antibody (PAb423: ref. 21) directed to the carboxy-terminal region of large-T as first

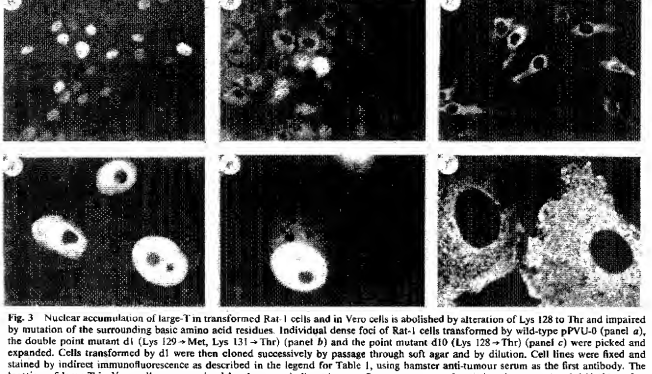

Fig. 3 Nuclear accumulation of large-T in transformed Rat-1 cells and in Vero cells is abolished by alteration of Lys 128 to Thr and impaired by mutation of the surrounding basic amino acid residues. Individual dense foci of Rat-1 cells transformed by wild-type pPVU-0 (panel *a*), the double point mutant d1 (Lys 129 → Met, Lys 131 → Thr) (panel *b*) and the point mutant d10 (Lys 128 → Thr) (panel *c*) were picked and expanded. Cells transformed by d1 were then cloned successively by passage through soft agar and by dilution. Cell lines were fixed and stained by indirect immunofluorescence as described in the legend for Table 1, using hamster anti-tumour serum as the first antibody. The location of large-T in Vero cells was examined by the same indirect immunofluorescence procedure using the monoclonal PAb423 as first antibody, 18 hr after manual microinjection[22] of about 200 copies of plasmid DNA into the nucleus of confluent cells growing on glass coverslips. The plasmids injected were wild-type pPVU-0 (panel *d*), d1 (panel *e*) and d10 (panel *f*). The fields shown in *d–f* are of confluent monolayers and include some cells that have not been microinjected and show no immunofluorescent staining.

antibody. This antibody, unlike anti-tumour serum, does not recognize SV40 small-t antigen which should be expressed in wild-type form by all mutant plasmids used in this study. The same cytoplasmic staining pattern was seen in Rat-1 cells transformed by d10 that were isolated as colonies growing in soft agar.

The exclusion of large-T from the nuclei of transformed Rat-1 cells consequent to mutation of Lys 128 is not observed for similar mutation of individual adjacent amino acid residues that form part of the five residue basic tract ([127]Lys-Lys-Lys-Arg-Lys[131]). Cells transformed by each of these mutants exhibit predominantly nuclear staining. In some cases (d2, X8, X46), the subcellular distribution of large-T appears indistinguishable from that of the wild-type protein. In other cases (X7, X18, d13, d27), a slight defect in nuclear accumulation of large-T is manifest by the presence of some cytoplasmic, in addition to nuclear, staining in the majority of cells derived from an individual focus. Furthermore, a minority of cells from the same focus show predominantly cytoplasmic staining.

The heterogeneous staining pattern (Fig. 3*b*) is even more obvious for the double mutant, d1 (Lys 129 → Met, Lys 131 → Thr). To date, the subcellular distribution of d1 large-T has been examined in 16 foci, derived from two independent transformation assays. In only one of these was a homogeneous staining pattern observed, with the mutant large-T apparently confined to the cytoplasm of the transformed cells. In all other cases, nuclear staining exceeded cytoplasmic staining in the majority of cells (60–95%); some cells clearly exhibited both nuclear and cytoplasmic staining, whereas in others large-T was apparently either wholly cytoplasmic or wholly nuclear. Cells derived from three of these foci were cloned further either by limiting dilution or by passage through soft agar. In each case, the heterogeneous staining pattern was preserved, indicating that clonally derived sibling cells express a large-T antigen that can, at any one time, be present predominantly in either cytoplasmic or nuclear compartments.

The variable distribution of the large T protein encoded by the mutants d1, d13, d27, X7 and X18 shows that mutations close to Lys 128 can modulate but do not prevent the accumulation of large-T in the nucleus. Although other interpretations are possible, the heterogeneous localization of large-T in cells transformed by these mutants may reflect a deficiency in the rate of nuclear accumulation less extreme than that observed for d10, which allows the visualization of intermediate stages in the nuclear accumulation of large-T, for example following the breakdown of the nuclear envelope at mitosis.

Transient expression

The wild-type plasmid, pPVU-0, and mutant derivatives were manually microinjected into the nucleus of semi-confluent Vero cells[22] and the location of the encoded large-T products assessed by indirect immunofluorescence 18 h post-injection. The staining patterns observed were virtually identical to those seen for transformed cells but were of much greater intensity, presumably due to the large number of gene copies available for transcription. Thus, large-T coded by wild-type pPVU-0 was exclusively nuclear (Fig. 3*d*) and that encoded by d10 was seen only in the cytoplasm (Fig. 3*f*). Both nuclear and cytoplasmic staining was observed for the mutants d1, d13, d27, X7 and X18 (Fig. 3*e*) but the heterogeneity in this staining pattern between individual cells in a population was not as marked as for transformed cell lines.

Replicative activity

As all mutations were introduced into a plasmid, pPVU-0 (ref. 23) that contains the viral origin of DNA replication as well as the entire large-T gene, the ability of the mutant large-T proteins to stimulate the replication of viral origin-bearing episomes in permissive cells could be tested directly by transfection of mutant plasmids into CV-1 cells and quantitation of intracellular plasmid DNA 72 h later[24]. In addition, the mutant large-T coding sequences from these plasmids were reconstructed into viral DNA and the viability of mutant virus tested by plaque assay[25]. The two assays gave similar results (Table 1), indicating that several of the mutants were unable to support viral DNA replication. Many other mutants[6] with lesions either side of those

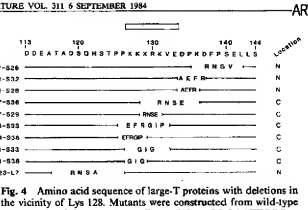

Fig. 4 Amino acid sequence of large-T proteins with deletions in the vicinity of Lys 128. Mutants were constructed from wild-type pPVU-0 containing one or two *Eco*RI linkers CGGAATTCCG at the site of a small deletion[9]. The consequent inferred alterations to the wild-type amino acid sequence are shown. Subcellular localization of mutant large-T antigens in transformed cells was determined as in Table 1 and Fig. 3. (+) (−) indicate the presence or absence, respectively, of nuclear immunofluorescence. Some of the large-T coded by T23-L7 was also detected in the cytoplasm. The box above the amino acid sequence of large-T indicates those residues which cannot be deleted without preventing nuclear localization of the mutants described here. Abbreviations used are: A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; V, valine.

described here also appear to be unable to stimulate viral DNA replication, indicating that this is a common phenotype of mutants from this region.

The viability of mutants altered at Lys 129 and Lys 131, respectively, implies that these (heterogeneously localized) mutant large-T antigens must be able to interact effectively with DNA sequences around the viral origin of replication. Direct measurement of DNA-binding activity *in vitro* by the McKay assay[26] suggests that the non-viable cytoplasmic (d10) and partially cytoplasmic (d1) mutants can also recognize SV40 origin DNA (E. Paucha, D.K. and A.E.S., manuscript in preparation), suggesting that the aberrant localization of these proteins is not related to their ability to interact with SV40 DNA.

Sequences around Lys 128

The data above show that alteration of Lys 128 and, to a lesser extent, the surrounding basic amino acid residues either modulates or abolishes the nuclear accumulation of large-T. The properties of a large number of single, double and triple point mutants with lesions either side of this region between amino acids 106 and 158, by contrast, do not exhibit any impairment in the nuclear accumulation of large-T[9]. The importance of Lys 128 to the localization of large-T is highlighted further by the recent observation of R. G. Lanford and J. S. Butel (personal communication) that the only lesion in the large-T coding sequence of the PARA-cT variant[27], isolated several years ago on the basis of its cytoplasmic phenotype, is the conversion of the codon for Lys 128 to that for Asn.

It is likely, therefore, that Lys 128 forms an essential part of some structural feature of the wild-type large-T protein that allows it either to be transported to or retained within the nucleus. The other possibility that the variant large-T proteins coded by d10 and PARA-cT have acquired an affinity for cytoplasmic components seems improbable, in that a novel specificity of binding is unlikely to be endowed by two different amino acid changes.

The context in which Lys 128 serves to localize large-T to the nucleus has been probed further by examining the properties of a number of insertion/deletion mutants with lesions in the immediate vicinity of Lys 128 (Fig. 4). These were made by standard procedures as part of a more comprehensive mutagenic analysis of large-T[9]. As all of the mutants shown in Fig. 4 were able to transform Rat-1 cells (Table 2), the subcellular distribution of the mutant large-T antigens could be examined in cells

Table 2 Transforming activity and subcellular localization of large-T proteins with deletions in the vicinity of Lys 128

	Net change in no. of amino acid residues	Nuclear large-T	Transformation of Rat-1 cells: No. of foci % wt 10 μg DNA	Transformation of Rat-1 cells: No. of foci % wt 1 μg DNA
L27-S26	−3	+	81	44
S1-S32	0	+	3, 20, 35*	30
S1-S28	+1	+	30	40
S17-S36	−4	−	66, 50	36
S17-S29	+1	−	38, 53	17
S14-S33	+1	−		
S14-S38	+2	−		
S11-S33	−3	−	72	75
S11-S38	0	−	110	35
T23-L7	−6	+	21	75

The transforming activity and subcellular locations of the mutant large-T proteins in transformed cells is given. These parameters were assayed as described in Table 1.

* These foci appeared 7, 3 and 6 days later than wild-type, respectively.

derived from transformed foci. In all cases where any part of the basic tract of amino acid residues (127–131) is interrupted, even if this does not include Lys 128, large-T is seen only in the cytoplasm of transformed cells (Fig. 4). However, the normal nuclear accumulation of large-T was not detectably impaired by the alteration of as many as four consecutive amino acid residues distal to Glu 133 and was only marginally affected by changing as many as ten residues proximal to Lys 127. Thus, if there exists a region of primary amino acid sequence around Lys 128 that directs large-T to the nucleus, it probably does not extend beyond Lys 127 or Glu 133. The tolerance of the nuclear large-T mutants L27-S26 and T23-L7 to variation in the length of polypeptide chain separating this sequence from the rest of the molecule perhaps also suggests that the function of Lys 128 is not dependent on a precisely structured three-dimensional configuration within the protein. The cytoplasmic location of several different insertion/deletion mutants supports the argument presented earlier that this phenotype does not result from a novel retention of large-T in the cytoplasm but rather from a defect in transport to or retention within the nucleus.

All of the cytoplasmic insertion/deletion mutants, like d10, transform Rat-1 cells according to dense focus formation in high serum with an efficiency and latent period similar to wild-type (Table 2).

Transformation by non-nuclear T

SV40 large-T is an unusual transforming protein in that it can transform primary cells as well as established cell lines[28] and gives rise to a gradation of transformed phenotypes[29]. In this study only the ability of large-T to transform an established cell line, Rat-1, in high serum was measured by the criterion of dense focus formation. None of the cytoplasmic large-T deletion or point mutants showed a significant deficiency in transforming activity relative to wild-type in this assay. Preliminary experiments indicate that the point mutant, d10, also transforms Rat-1 cells, by the criterion of colony formation in soft agar at high serum concentration with an efficiency similar to wild-type. However, any implication that large-T need not be nuclear to transform, must be tempered by two considerations. First, although large-T is seen by immunofluorescence only in the cytoplasm of cells transformed by d10 or by any of several deletion mutants, some large-T may nevertheless be present in the nucleus but remain undetected, either because of its very low abundance or because it is in a form that is not accessible to the antibodies used for its detection. Second, it is possible that differences in the transforming activity of nuclear and non-nuclear large-T species may be revealed by more stringent assays. By using such assays, R. G. Lanford and J. S. Butel

concluded that the cytoplasmic large-T encoded by the PARA-cT mutant induces a partially transformed phenotype in hamster cells, although it can induce colony formation in semi-solid medium and tumour formation *in vivo*[30]. In view of the notion that the transformation of primary cells involves the combined action of a nucelar immortalizing gene product (such as polyoma virus large-T, adenovirus E1a protein, the *myc* gene product) and a membrane-associated protein[31] (such as the *ras* gene product, polyoma virus middle-T), it might be predicted that a non-nuclear large-T would be unable to immortalize or transform primary cells, whereas the plasma-membrane associated form of large-T[5] may suffice to transform established cell lines. These predictions are being tested using the cytoplasmic mutants described here.

Mechanisms of nuclear localization

The nuclear envelope differs from other cellular membranes in that it allows the free passage of virtually all proteins below a certain size (corresponding to approximately 18,000 M_r) in addition to some larger proteins[1,2,4]. Thus, for any given protein it is not clear whether nuclear accumulation results from a specific mechanism that either facilitates or drives transport across the envelope or whether the protein is free to diffuse, for instance through nuclear pores and is then retained in the nucleus by virtue of an affinity for elements that are themselves confined to the nucleus. This distinction has not been made for SV40 large-T, though the large size of the monomer may favour a specific transport mechanism.

Received 24 April; accepted 3 July 1984.

1. Toore, J. *Molecular Biology of Tumor Viruses* Part 2, 2nd edn (Cold Spring Harbor Laboratory, New York, 1981).
2. Tegtmeyer, P. *J. Virol.* **10**, 591–598 (1972).
3. Pope, J. H. & Rowe, W. P. *J. exp. Med.* **120**, 121–128 (1964).
4. Sanna, M. & Butel, J. S. *Virology* **120**, 1–17 (1982).
5. Klockmann, U. & Deppert, W. *EMBO J.* **2**, 1151–1157 (1983).
6. Tjian, R. *Cell* **13**, 165–179 (1978).
7. Shortle, D. R., Margolskee, R. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6128–6131 (1979).
8. Rio, D., Robbins, A., Myers, R. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5706–5710 (1980).
9. Kalderon, D. & Smith, A. E. *Virology* (in the press).
10. Prives, C., Beck, Y. & Shure, H. *J. Virol.* **33**, 689–696 (1980).
11. Chanady, F., Harvey, R. & Smith, A. E. *J. Virol.* **44**, 54–66 (1982).
12. Clark, C., Peden, K., Pipas, J. M., Nathans, D. & Tjian, R. *Molec. cell. Biol.* **3**, 220–228 (1983).
13. Morrison, B., Kress, M., Khoury, G. & Jay, G. *J. Virol.* **47**, 106–114 (1983).
14. Scheidtmann, K. H., Echle, B. & Walter, G. *J. Virol.* **44**, 116–133 (1981).
15. Hutchinson, C. A. III *et al. J. biol. Chem.* **253**, 6551–6560 (1978).
16. Smith, M. & Gillam, S. in *Genetic Engineering, Principles and Methods*, Vol. 3 (eds Setlow, J. K. & Hollaender, A.) (Plenum, New York, 1981).
17. Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
18. Wallace, R. B. *et al. Nucleic Acids Res.* **6**, 3543–3557 (1979).
19. Wallace, R. B. *et al. Gene* **16**, 21–26 (1981).
20. Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456–467 (1973).
21. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861–869 (1981).
22. Richardson, W. D. & Westphal, H. *Cell* **37**, 133–141 (1981).
23. Kalderon, D., Oosters, B. A., Ely, B. K. & Smith, A. E. *Nucleic Acids Res.* **10**, 5161–5171 (1982).
24. Myers, R. M. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6491–6495 (1980).
25. Dulbecco, R. & Freeman, G. *Virology* **8**, 396–397 (1959).
26. Mellor, R. D. G. *J. molec. Biol.* **145**, 471–488 (1981).
27. Rapp, F., Paulucci, S. & Butel, J. S. *J. Virol.* **4**, 626–634 (1969).
28. Pollack, R., Risser, R., Conlon, S. & Rifkin, D. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4792–4796 (1974).
29. Risser, R. & Pollack, R. *Virology* **59**, 477–489 (1974).
30. Lanford, R. G. & Butel, J. S. *Virology* **110**, 147–158 (1981).
31. Land, H., Parada, L. F. & Weinberg, R. A. *Science* **222**, 771–778 (1983).
32. Paine, P. L. & Feldherr, C. M. *Expl Cell Res.* **74**, 81–98 (1972).
33. Paine, P. L., Moore, L. C. & Horowitz, S. B. *Nature* **254**, 109–114 (1975).
34. De Robertis, E. M. *Cell* **32**, 1021–1025 (1983).
35. Dingwall, C., Sharrick, S. V. & Laskey, R. A. *Cell* **30**, 449–458 (1982).
36. Hall, M. N., Hereford, L. & Herskowitz, I. *Cell* **36**, 1057–1065 (1984).
37. Prentki, P., Karch, F., Iida, S. & Meyer, J. *Gene* **14**, 289–299 (1981).
38. Edge, M. D. *et al. Nature* **292**, 756–761 (1981).

The accumulation of nucleoplasmin in the nucleus of *Xenopus laevis* oocytes[35] is the only case for which there is good evidence for the involvement of a nuclear transport system. Proteolytic digestion of the native pentameric protein has shown that a 16,000 M_r fragment derived from a single 33,000 M_r monomeric unit is sufficient to allow the nuclear accumulation of partial proteolytic cleavage products as large as 101,000 M_r, and is therefore presumed also to specify the nuclear location of the native protein.

Whatever the mechanism of nuclear accumulation, the evidence presented here heralds the possibility that a very restricted region of the SV40 large-T amino acid sequence around Lys 128 may be sufficient to ensure the nuclear location of this protein. Further experiments are in progress to determine more precisely to what extent this putative signal is independent of the rest of the large-T molecule and whether the signal can act in the context of other proteins that are normally cytoplasmic so as to direct them to the nucleus. The availability of the d10 mutant will be invaluable as a control for such studies. When this manuscript was in preparation, M. N. Hall *et al.* reported that restricted sequences derived from the yeast Mat α2 gene product can direct another protein to the nucleus in yeast[36].

We thank Robert Lanford, Janet Butel and Lynna Hereford for communication of results prior to publication, Ed Harlow for monoclonal antibody PAb423, Roger Morris for the gift of rhodamine-isothiocyanate-conjugated rabbit anti-mouse antibody, Joe Brock for help with illustrations and Lydia Pearson for typing the manuscript.

LETTERS TO NATURE

The boundary of the Solar System

R. Smoluchowski

University of Texas, Austin, Texas 78712, USA

M. Torbett

Carnegie Institution, Department of Terrestrial Magnetism, Washington DC 20005, USA

The shape of the boundary of the Solar System, defined as the surface within which the gravitational attraction of the Sun, rather than that of the rest of the Galaxy, controls the orbital motion of bodies such as planets and comets, has been determined. Outside of this surface, the dominant factors are the radial tides due to the galactic centre and, especially, the vertical tides caused by the galactic disk. Orbits which are direct with respect to the galactic plane, have a boundary which differs from that for retrograde orbits, both being 10–20% oblate and both larger than the, usually assumed, present Oort cloud. The surface may have been the boundary of the early cloud of comets which was later reduced by the passages of stars and molecular clouds.

Losses of bodies from the Solar System are caused by: sporadic perturbations by passing stars or giant molecular clouds, which occur at a rate of several per 10^6 or 10^7 yr respectively, perturbations by giant planets and the steady gravitational perturbation by the galaxy. The first two kinds of perturbations have been extensively studied[1–4]. Here we describe the results of a three-dimensional stability study (inertial coordinates) of a system consisting of a comet and the Sun rotating[5] around the galactic centre [mass 1.3×10^{11} $M_\odot$, distance 8.2 kpc (ref. 6)] and periodically traversing the nearly harmonic field of the galactic mid-plane (GMP)[7]. A two-dimensional model of a greatly simplified version of this problem has been studied previously[3] and a corresponding three-dimensional Hill surface is also known[8]. To get an indication of the extent and structure of the boundary of stability[9] paths of comets were integrated numerically for various initial parameters using Cowell's method and fourth-order Runge–Kutta integration. Near the boundary of stability the orbit of a comet is non-keplerian, its semimajor axis a, average inclination i, eccentricity e and the direction of perihelion $\bar{\omega}$ are highly variable. Thus, for an initial i, the values of e and a were so chosen as to obtain an orbit which was bound for the age of the Solar System. The value of $\bar{\omega}$ had essentially no effect on the results.